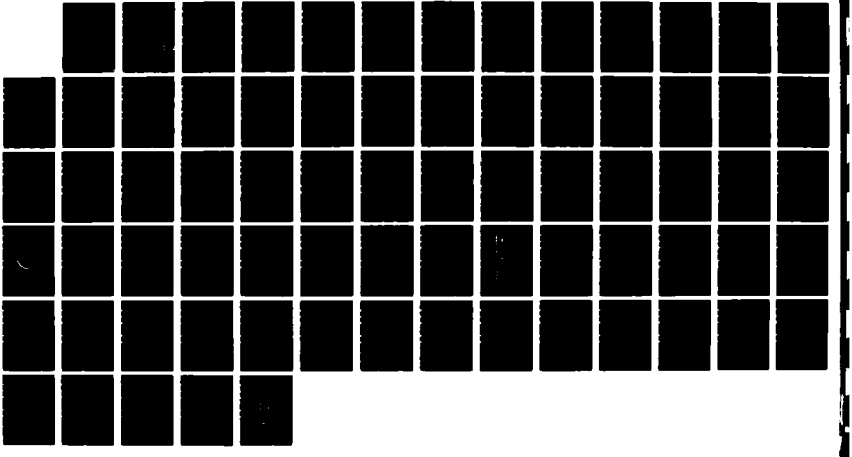
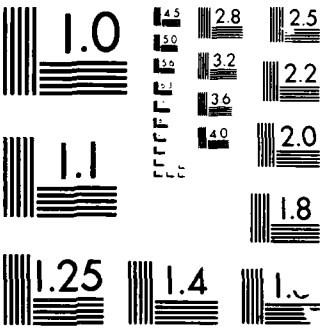
AD-A191 993 MOLECULAR MECHANISMS OF CYTOPATHOGENICITY OF PRIMATE 1/1
LYMPHOTROPIC RETROVI (U) BIOTECH RESEARCH LABS INC
ROCKVILLE MD M M MANAK ET AL 28 OCT 87
UNCLASSIFIED DAMD17-86-C-6287 F/G 6/13 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

ORIGINAL

AD_____

AD-A191 993

MOLECULAR MECHANISMS OF CYTOPATHOGENICITY OF PRIMATE LYMPHOTROPIC RETROVIRUSES: RELEVANCE TO TREATMENT AND VACCINE FOR AIDS

Annual Report

Covering the Period 9/29/86 to 9/28/87

by

Mark M. Manak and Linda L. Jagodzinski

October 28, 1987

Supported by

US Army Medical Research and Development Command
Fort Detrick, Frederick, Maryland 21701-5012

Contract No. DAMD17-86-C-6287

DTIC
ELECTE
FEB 19 1988
S D
H

Biotech Research Laboratories, Inc.
Rockville, Maryland 20850

Approved for public release; distribution is unlimited.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

88 2 18 032

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Biotech Research Labs., Inc. / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): Rockville, Maryland 20850	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAMD17-86-C-6287
8c. ADDRESS (City, State, and ZIP Code): Fort Detrick, Frederick, Maryland 21701-5012	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 623105 / PROJECT NO. 3M2 623105H29 / TASK NO. AC / WORK UNIT ACCESSION NO. 016

11. TITLE (Include Security Classification)
Molecular Mechanisms of Cytopathogenicity of Primate Lymphotropic Retroviruses: Relevance to Treatment and Vaccine for AIDS

12. PERSONAL AUTHOR(S)
Manak, Mark M. and Linda L. Jagodzinski

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 9/29/86 TO 9/28/87	1987 October 28	66

16. SUPPLEMENTARY NOTATION
Simian T-Cell Lymphotropic Virus Type III (African Green Monkey)

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	HIV-1, Lymphotropic Viruses, Cytopathogenicity, DNA Sequence, Gene Function
06	03		Human Immunodeficiency Virus Type 1
06	13		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The molecular basis of infectivity, cytopathogenicity and genomic activation of (HIV-1) was investigated by generating deletion mutations in sor, 3'orf, 5' region, and LTR. These studies show that the sor gene is crucial for the generation of infectious virus. Analysis of 3'orf indicate that deletions in the amino terminus (which overlaps the carboxy terminus of gp41) result in significant reduction in viral propagation. Studies with the LTR sequence have delineated an enhancer activity, an sp-1 bending site, and a tat response region (TAR).

Genomic clones containing the integrated provirus of (STLV-III$_{AGM}$) and HIV-2$_{NIH-Z}$ have been isolated. Their nucleic acid sequence has been determined and subjected to computer analysis. A Comparison of the nucleotide sequences of these two new viral isolates with HIV-1 indicates that these new retroviruses are structurally similar to HIV-1 and belong to the same class of viruses as HIV-1. STLV-III and HIV-2$_{NIH-Z}$ exhibit a greater homology among themselves than with HIV-1. A study of biologically active forms of these viruses will aid in understanding why these new viruses are non-pathogenic.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Mary Frances Bostian	301-663-7325	SGRD-RMI-S

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

TABLE OF CONTENTS

QUALITY INSPECTED

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

TABLE OF CONTENTS (Cont'd) PAGE

LIST OF FIGURES AND TABLES

FOREWORD

In conducting the research described in this report, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals", prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (DHEW Publication No. (NIH) 86-23, Revised 1985).

The investigator(s) have abided by the National Institutes of Health Guidelines for Research Involving Recombinant DNA Molecules (April 1982) and the Administrative Practices Supplements.

SUMMARY

The molecular basis of infectivity, cytopathogenicity and genomic activation of HIV-1 was investigated. Site directed mutagenesis and restriction endonuclease cleavage followed by Bal 31 digestion was used to generate deletion mutations in sor, 3'orf, the 5' packaging region and LTR. The biological activity of the mutations was assessed by transfecting the mutants into permissive cells and examining the cells for expression of viral antigens, reverse transcriptase, viral particles, and infectivity. Analysis of these mutants to date show that the sor gene is not absolutely required for HIV virion formation but influences viral transmission in vitro and is crucial for the generation of infectious virus. Analysis of 3'orf activity indicated that deletions in the amino terminus (which also results in truncation of the carboxyl terminus of gp41) result in significant reduction in viral propagation/infectivity. Studies with the LTR sequence have delineated an enhancer activity on Sp-1 binding site and the tat response region (TAR).

Within the last two years several new retroviral isolates have been obtained which exhibit a high degree of homology with the human immunodeficiency virus type 1 (HIV-1). The discoverey of a simian virus ($STLV\text{-}III_{AGM}$) and a second class of human retroviruses (HIV-2) which are highly related to each other and distinctly related to HIV-1 has led researchers to hypothesize that these viruses all diverged from a common ancestor. Furthermore, these new viral isolates appear to be non-pathogenic. In order to determine the differences that exist between these viruses and HIV-1, we have generated lambda phage clones of the integrated provirus for both $STLV\text{-}III_{AGM}$ and $HIV\text{-}2_{NIH}$. The nucleotide sequences of these genomic clones have been determined and compared with HIV-1 and each other. The deduced amino acid sequences have also been compared with that of HIV-1. Structurally, these viral isolates are very similar to HIV-1 except for a premature stop codon in the envelope gene and an extra reading frame designated X. The homology of the nucleic acid sequence between STLV-III and HIV-1 is 55% overall, whereas, STLV-III and $HIV\text{-}2_{rod}$ exhibit an overall homology of 75%. This sequence data indicates that these isolates do belong to the same group of retroviruses as HIV-1. A comparison of the deduced amino acid sequence of the genes encoded for in the provirus confirms a conservation of the protein structure.

BODY OF THE REPORT

A. DEVELOPMENT OF DEFECTIVE HIV PARTICLES

1. Statement of Problem

A major effort in the fight to stop the rapid spread of AIDS has been devoted to the development of an effective vaccine against the human immunodeficiency virus-1 (HIV-1), the etiological agent of the disease. The classical approach, production of an attenuated virus vaccine, is not appropriate in the case of HIV-1 due to the extremely high lethality of the virus. Even very low levels of biologically active virus may be capable of initiating productive infection leading to the disease. Since the virus attacks T4 cells, which play a central role in regulating the immune system, an effective immune response may not be launched in time to prevent infection. Furthermore, the virus has the capacity to reside for long periods of time in an integrated form (provirus) in cells in the absence of detectable symptoms. This latent virus may subsequently be activated to initiate the disease. Even inactivated or killed virus may not be suitable for vaccines since the viral genome can integrate into that of the host cells and thereby transform them.

One approach that may still be available for production of a viral vaccine would be to use recombinant DNA technology to generate deletion mutants of HIV-1 that are not infections and can not produce cytopathic effects. Such deletions should be made in genes which regulate virual replication and cytopathogenicity but do not interrupt the viral structural genes which would retain the antigenic characteristics of the virus. A clear understanding of the molecular basis of cytopathogenicity and regulation of infectivity is essential in designing appropriate mutant viruses and to assure their "safety". This information is also critical in designing therapeutic protocols to block the devastating effect of the virus on the immune system.

2. Background

The study of the molecular basis of HIV-1 infection has been made possible by the establishment of an *in vitro* model for AIDS in which the virus can be propagated in PHA-stimulated lymphocytes or in a variety of lymphoid cell lines (1). Further progress has been possible with the use of molecularly cloned proviral DNA which could be transfected into these cultured cells. This DNA can then direct the synthesis of infectious virus particles which exhibit biological properties indistinguishable from virus obtained by cocultivation of permissive cells with the peripheral blood mononuclear cells of AIDS patients (2). The availability of

a cloned genome and an *in vitro* system of cultivating virus has permitted extensive analysis of the function of the viral genome and the viral proteins for which it codes.

The HIV genome consists of a 9.7 kilobase RNA molecule, in which eight genes have been identified to date. The three structural genes, *gag*, *pol*, and *env* are common to all retroviruses and constitute the largest open reading frames of the genome. In addition, the virus contains *tat-III* and *art/trs* which represent regulatory genes for virus replication, and three genes of undefined function *sor*, *3' orf*, and *R* (3).

In the course of this contract, efforts have been concentrated on elucidating the function of the nonstructural genes, particularly *tat-III*, *sor* and *3' orf*. *Tat-III* will be discussed more fully in Section B of this report. The *sor* gene (for short open reading frame) of HIV (also called *Q*, *P'*, *Orf-1*, and *Orf A*) lies between the *pol* and *tat* genes overlapping at its 5' end with *pol* (4). It is an open reading frame of 609 nucleotides in size and encodes a protein of 23 Kd (5). This gene appears to be conserved among all HIV isolates, in the distantly related SIV, and in other retroviruses suggesting that this gene may be functionally significant. Preliminary studies with *sor* deletion mutants of HIV, however, suggested that it is not essential for virus replication or cytopathic effects. Although small reductions in infectivity were noted, the significance of this observation was not well understood.

The *3' orf* gene (for 3' open reading frame) has the coding capacity of a protein of about 27 Kd and spans the region between *env* and 3' LTR overlapping both genes. Preliminary studies in Dr. Robert Gallo's lab with deletion mutants in *3'orf* suggested that this gene may be important in virus induced cell killing.

A non cytopathic HTLV-III clone, designated X10-1, was generated from the cytopathic genome of pHXB2D by removing a 200 base pair segment around the XhoI site in the 3' region of the provirus. This deletion removed the coding capacity of the last 5 amino acids of the transmembrane portion of the envelope, together with the first 60 amino acids of *3' orf*, and shifted the remaining codons of this gene out of frame. This clone generated replicating virus when transfected into normal T-lymphocytes, and T-lymphoblastoid cell lines. This virus is no longer cytopathic for normal T-cells, and does not kill ATH5 cells which rapidly and reproducibly succumb to HTLV-III mediated killing by conventional virus isolates. Interestingly, clones which contain less extensive deletions (55-109 base pairs), restricted to *3' orf*, generate

virus which is extremely cytopathic. These data argue that cell killing may be mediated by the carboxy-portion of HTLV-III envelope or by the first 22 amino acids of 3' orf. Deletions of the 3'orf gene may therefore be expected to lead to the production on non-cytopathic virus particles.

An alternative approach to producing viral particles which are non replicating is to generate "packaging" defective mutants which do not encapsidate the RNA and therefore produce only "empty" particles. Although much is known about retroviral replication and the processing of gag, pol, and env products in infected cells, the precise mechanics of virus particle formation and the mechanism by which the viral genome is prefentially packaged into virions remains unclear. Presumably, out of the array of viral and cellular mRNA's, ribosomal, and transfer RNA's found in infected cells, this mechanism permits the virus assembly apparatus to select and incorporate its own genomic viral RNA.

Studies in the avian and murine retrovirus system have suggested that virus particle formation can occur in the absence of genomic RNA (7) and sequences that intervene at the 5' LTR and gag are crucial for virus specific packaging (8, 9, 10). Deletions in the sequences interviewing at 5' LTR and gag regions of HIV-1, therefore, may be expected to give rise to "empty" (non-viral RNA containing particles, not containing viral RNA).

3. Rationale

The specific goals of this portion of the contract are to produce defective HIV-1 particles which contain the viral structural proteins, but are non-infectious and non-cytopathic. To accomplish this, it is first necessary to clearly define the biological activity of the non-structural proteins and the role they play in infectivity and cytopathogenicity. One approach to the study of viral function is to introduce specific mutations in a biologically active virus genome, and examine the consequences of such alteration upon subsequent transfection into recipient cells. Specific deletions or alterations of the viral genome can be accomplished by site directed mutagenisis or by the combined actions of exo- and endo-nucleases.

Three portions of the viral genome were selected for detailed study: the sor and the 3'orf reading frames and the 5' end of the genome involved in packaging recognition. As discussed in the Background section, mutations in sor may be expected to generate particles which can replicate and induce cpe, but which have reduced infectivity. Deletions in the 5' end of gp41 are of special interest in this regard.

Mutations in 3'orf may be expected to be defective in cell killing. Finally packaging defective mutants may produce "empty" particles incapable of replication or inducing cytopathic effect. By introducing a series of mutations at specific sites within these regions, the biological activity of resulting clones can be subsequently determined in transfected cells. Assays for replication, packaging, and cytopathic effect in transfected cells need be designed to properly characterize the nature of each mutation defect. Using this type of strategy, it should be possible to precisely localize the genetic sequence and function responsible for infectivity, cytopathogenicity and packaging.

4. Experimental Methods

Sor Deletion Mutants - A series of sor mutants were generated from pHXB2gpt (11) by removal of an EcoRI site in the polylinker of the vector. Mutant S was prepared by removal of the sequences between the NdeI and NcoI restriction sites (nucleotides 4707 to 5259) and religation, additional mutants were prepared by site specific mutations using oligonucleotide-directed mutagenesis. These mutations introduced translational stop codons at various points in the sor frame downstream of the sor/pol overlap (Fig. 1). To construct these mutants the EcoRI to EcoRI fragment (nucleotides 4230 to 5322) of the HTLV-III genome of BH10 was subcloned into an M13 phage vector, and mutagenesis performed as described previously (12). Sequences in this region of BH10 differ from those of HXB2 only at nucleotide position 4506 which results in a proline to serine substitution in pol. Mutation 6.9 was introduced using the 25-mer (GGATGAGGGCTTTCTTAGTGATGCT), converting the tyrosine codon (TAT) at residue 55, into a stop codon (TAA). In a similar approach was used to replace the serine codon (TCA) at position 42 was replaced with a stop codon (TAA) in clone 3.3, and to change the glutamine codon (GAA) at residue 100 to a stop codon (TAA) in clone 153. Following confirmation by DNA sequencing the mutated fragments were subcloned into the original proviral clone.

3'orf Mutations - A series of deletion mutations in the 3'orf region were constructed from pHXB2-D in collaboration with Lee Ratner. Purified phage DNA was cut with XhoI and subjected to Bal31 digestion. Aliquots were removed at various times of digestion and religated to generate a series of deletions mutations about the XhoI site (Fig. 2). Deletions ranging in size from approximately 150-500 nucleotides and extended into the gp41 region were selected for further study.

Packaging Defective Mutants - Packaging defective mutants were made by producing deletions in the region between the 5'LTR and gag coding sequences of the plasmid clone pHXB2gpt.

This plasmid consists of a biologically active HTLV-III genome combined with the E. coli xanthine guanine phosphonoribosyl phosphatase gene (gpt) and contains a unique BssHII site in the sequences that intervene the 5' LTR and the beginning of the gag gene. A series of deletion mutants about the BssHII site were generated using the Bal 31 digestion-regulation approach (Fig. 3). Mutants pHXB2^{3} and pHXB2^{11} are deletions of 51 and 57 nucleotides in positions 224-227 and 224-283, respectively. Additional mutants were obtained by site directed mutagenesis. The mutant X10-1^{25} is deleted of 9 nucleotides between positions 310-320. The mutant #293 contains deletions between 293 and 328. In all cases, the splice donor site (position 287) is preserved.

Characterization of Deletion Mutants - Purified DNA from the various deletion mutants produced as described above was transfected into permissive human lymphoid cells (H9, Molt 4, or PHA-stimulated blood mononuclear leukocytes) or to the monkey kidney fibroblastoid line Cos-1 using the calcium phosphate precipitation technique (13). In these experiments 10 ug of cesium chloride gradient purified plasmid DNA was introduced into 1 x 10^{6} cells. Since the plasmid clones contain the gpt gene (which allow recipient cells to grow in the presence of microphinolic acid), stably transfected cell lines containing the transfected genomes could be selected by growth in appropriate media (HAT). The transfected cultures were monitored at approximately weekly intervals for the appearance of HIV-1 gag and env by indirect immunofluorescence with antibodies to p17, p24, and gp41. Expression of the functional pol gene product was determined by reverse transcriptase assay. The presence of viral particles was examined by electron microscopy.

5. Results

Deletions of Sor

DNA from each of the sor deletion mutants was transfected into permissive lymphoid cells (H9, Molt4 and PHA-stimulated blood mononuclear leukocytes), and the transfected cultures were monitored at approximately weekly intervals for HIV-1 expression. Molt 4 cells, normal T-cells or H9 cells transfected with the deletion sor mutants consistently failed to express virus as detected by immunofluorescence, reverse transcriptase assays or electron microscopy. Southern blotting analyses on these samples showed transient uptake of plasmid DNA but no detectable proviral sequences in long-term cultures. In contrast, cells transfected with pHXB2gpt reproducibly and rapidly yielded virus producing cells.

The ability of the sor mutants to produce virions was also

examined by performing a series of transfections using the SV40 transformed cell line Cos-1 as a target. These experiments were aimed to amplify virus production by exploiting the capacity of the Cos-1 cells to promote episomal replication of plasmids carrying the SV40 origin of replication (including plasmids derived from pHXB2gpt) in transient assays. The transfected Cos-1 cells were then cocultivated with Molt 3 cells which are permissive for HIV-1 infection.

As shown in Table 1, virus particles morphologically similar to wild type were recovered from Cos-1 cell cultures transfected with sor mutants 6.9, 3.3, 153, and S. The level of virus production (both extracellular and budding virions) and transactivation potential was indistinguishable from that seen in pHXB2-gpt transfected cultures. However, supernatants removed from sor mutant cultures (containing cell free virus) failed to infect H9 cells in repeated attempts. These results indicate that the sor gene of HIV is not critical for the production of normal levels of morphologically intact virus particles, but mutant viruses are limited in their capacity to establish stable infection.

Virus derived from transfection of Cos-1 cells (OKT4-, African Green Monkey Cells) with sor mutant proviral DNA's, was resistant to transmission to OKT4+ 'susceptible' cells under cell-free conditions, and was transmitted poorly by coculture. In contrast, virus derived from clones with an intact sor frame was readily propagated by either approach. Normal amounts of gag, env and pol derived proteins were produced by all four mutant genomes and assays performed in both lymphoid and non-lymphoid cells indicated that their trans-activating capacity was intact and comparable with wild type. These data show that the sor gene, although not absolutely required in HIV virion formation, influences virus transmission in vitro and is crucial in the efficient generation of infectious virus. This data also suggest that sor influences virus replication at a novel, post-translational stage and its action is independent of the regulatory genes tat and art/trs.

Deletions in 3'orf

A total of 36 HIV-1 mutants with deletions in the 3' region of the genome have been constructed and characterized. Of these, clones 369, 468, 372 and 429 were generated by site directed mutagenesis, while the remainder were obtained by Bal 31 digestion at the XhoI site. The results of sequence analysis across the deleted regions of these plasmid constructs have now been completed and are summarized in Table 2. Briefly, these mutants fall into 4 categories; those with deletions solely within the 3'orf regions (clones 230, 330),

clones with alterations in 3'orf and gp41 env (clones 329 through 362, Table 2) and clones with an additional alterations in the neighboring 3' regions (LTR and polypurine tract clones 358-194) or 5' regions (tat-III and trs clones 360-327). The resulting alterations in the carboxyl end of gp41 is indicated by the number of amino acids deleted and the number of additional amino acids added in frame.

Analysis of the biological properties of these HIV-1 mutants has been carried out at two levels. First, the ability of mutant genomes to give rise to intact virus particles with reverse transcriptase activity, was analysed by transfecting each plasmid into Cos-1 cells and examining the culture supernatant for virion production by EM and RT assays. The data presented in Table 2 shows that all constructs were replication competent except for those containing deletions in the tat-III or trs genes (clones 306, 327). This is consistent with previous reports that the tat and trs genes are essential for virus replication. A second analysis was carried out to determine the ability of env mutants to be propagated in culture. This was done by co-culturing transfected Cos-1 cells with OKT4 positive recipient cells (either Molt 3 or H9 cells). Although the analysis is only half completed preliminary results suggest that: (1) relatively small changes in the carboxyl region of gp41 env, diminish the propagation/infectivity of virus; (2) these effects are more markedly seen when Molt 3 cells are used as recipients rather than H9 cells; and (3) deletions which enter the LTR region results in virus which cannot be efficiently propagated.

Packaging Mutants

Studies of the biological properties of the packaging mutants have not been completed. However, the initial analysis has shown that clones $pHXB2^{3}$ and $pHXB2^{11}$ generate morphologically normal virus particles upon transfection into the Cos-1 cell line. Furthermore, these particles are infectious (and therefore, presumably contain genomic RNA) and can easily be transmitted to recipient OKT4+ cells. These data argue that sequences between 224-283 are unlikely to be important in the process of viral packaging. Mutant $X10\text{-}1^{25}$, and other clones bearing more extensive deletions downstream of the splice donor are currently under study.

6. Discussion

Mutations in Sor

The studies described in this report have shown that construction and analysis of deletion mutations is a very powerful approach to elucidating gene function. We have

shown that removal or truncation of sor results in virus progeny that have a much reduced (>100-fold) capacity to infect. Furthermore sor mutant viruses were transmitted less well under co-culture conditions (in which cell to cell transmission is likely to be important). Interestingly, the effects of truncating sor (in the case of mutants 3.3, 6.9, 153) were similar to its complete removal (in the case of mutant S) suggesting that the carboxy-terminal portion of sor (downstream from residue 100) may include a functional domain. However, the possibility that sor is non-functional because it is deprived of critical elements in the carboxy-terminal of the protein (necessary for the correct folding or processing) is not excluded. Since the level of viral RNA, proteins and viral particles produced by sor defective genomes could not be distinguished from that of wild type, we suspect that sor exerts its effects at a post-translational level. This novel regulatory mechanism, mediated by sor, could involve late events in virus maturation. These findings substantiate the complex nature of the HIV genome, underscoring the sophisticated transcriptional, post-transcriptional and post-translational controls which operate to regulate virus expression in infected cells. While the mechanism of how sor enhances virus propagation is not yet understood, several distinct mechanisms could be postulated and are currently being investigated. It is possible that sor is a structural component of the virion particle which acts as a 'second envelope', required for efficient transmission. Since it is difficult to detect sor in as large amounts as gp120 and gp41 (in either infected cells or virions) additional studies are necessary to evaluate this likelihood. Alternatively, the sor gene might be involved in stabilizing or processing envelope so that assembly of infectious virus is increased, or in potentiating the cellular environment in which viral replication occurs.

Mutation in 3'orf

Studies with deletions in the 3'orf gene have shown that this gene is not essential for virus particle formation or reverse transcriptase activity. However, even small changes in the carboxyl region of gp41 env (which overlaps the 3'orf) greatly diminish viral infectivity. Thus virus produced in Cos-1 cells which amplify the gene expression, transmit poorly to H9 cells and even less well or not at all to Molt 3 cells. The host cell itself plays a large role in this decrease since these effects are more pronounced when Molt 3 cells are used as recipients rather than H9 cells.

The cytopathic properties of the 3'orf mutants is currently being evaluated using the sensitive ATH8 cells or PHA stimulated normal T-cells as targets. Immunoprecipitation studies are

also being performed to visualize the truncation-elongation of env derived gp41, gp120, and gp160 (precursor) proteins, created by the mutations. Finally, the interactions of gp41 and HLA class I molecules will be investigated to try and see whether the diminished propagation of gp41 mutants in Molt 3 cells might result from their failure to interact with class I determinants as has been suggested from studies by Fulvia Veronese (Litton Bionetics) and collaborators (diMarzo-Veronese, unpublished).

Packaging Mutants

The preliminary characterization of the 5' deletion mutants of sequences between 223 and 284 have failed to detect alterations in virus morphology or infectivity. Additional characterization of these mutants and of those containing more extensive deletions are currently in progress. The specific infectivity (infectious centers/ug) will be determined on Cos-1 cells. Transfections of various lymphoid cells will be attempted to generate lines of infected cells.

These stable cell lines, will be characterized in terms of viral protein expression (Western blotting, immuno-precipitation, and immunofluorescence), expression of viral RNA species (Northern blotting) and the capacity to generate virion particles. To determine whether these particles contain virus specific RNA (i.e. are correctly 'packaged'), Northern blotting of sucrose banded 'purified' virus particles will be attempted using a BH10 insert labeled with ^{32}P as a probe, and purified "wild type" virus as a positive control. The amount of viral RNA present per virion particle will be quantitated to verify whether 'empty' particles are in fact produced.

B. MOLECULAR STUDIES OF TAT-III ACTIVITY

1. Statement of Problem

The level of HIV gene expression in infected cells is subject to a complex set of virus encoded regulatory elements. Infection by HIV-1 *in vivo* may be associated with an asymptomatic interval frequently lasting from months to years before the development of actual disease (14). Such a latent state of infection can also be duplicated *in vitro*, where HIV-1 infected CD4+ cell lines can be maintained for long periods of time in the absence of detectable virus production (15). Events that trigger the transition from low level or latent infection to productive viral replication remain poorly defined. Immune activation of infected T-cells can greatly stimulate HIV replication, apparently through effects on the regulatory elements within the long terminal repeat. An understanding of the mechanism of HIV-1 gene activation is essential for designing therapeutic approaches to control the onset of the disease in HIV-1 antibody positive individuals. This information will also have major implications in the design of 'safe' vaccines in which the genome will not be subject to rapid activation.

2. Background

Expression of the HIV-1 genome is tightly regulated by the activity of several viral genes and regulatory sequences. The *tat-III* gene located between the *sor* and *env* genes has been reported to enhance transcription of genes linked to the long terminal repeats LTR's in infected cells (*tat* response), and presumably plays a central role in HIV-1 gene activation (16). The gene is expressed as a 1.9-2.0 kb mRNA generated by two splicing events and has many similarities to the corresponding *tat* gene in HTLV-I and HTLV-II (17). However, whereas the *tat* I and II proteins are reported to enhance transcription, the *tat* III gene product was shown to enhance expression of protein at a posttranscriptional level (18). Another gene designated *art* or *trs*, possibly expressed from the same mRNA which encodes the *tat*-III gene, also regulates HTLV-III virus expression posttranscriptionally, most likely by regulating of the accumulation of genomic and spliced viral mRNA (19).

The level of gene expression has been reported to be dependent on functional elements within the LTR sequence of the virus, including; (i) a negative regulatory element (NRE), (ii) an enhancer upstream of the promoter or TATA box, and (iii) *tat* response region downstream of the promoter at -17 to +80 (20). The *tat* response region allows greatly increased expression of linked genes in HTLV-III infected cells relative

to uninfected cells even if placed in a similar position downstream of heterologous promoters (20). The HTLV-III-LTR was also reported to contain three Sp1 binding sites as demonstrated by point mutagenesis and protection experiments (21). However, the functional discrimination of these sites was only partially dissected. The HIV-1 LTR contains both *cis* and *trans* regulating elements which have not been precisely localized. Much work remains to be carried out in defining the function of the regulatory elements and their interactions with the viral genome.

3. Rationale

A series of deletions within the HTLV-III LTR was introduced to localize the *cis* acting elements upstream of the promoter within U3 and to localize the *trans* acting response sequences in the R region. To study the resulting LTR function of these alterations, these constructs were linked to the chloramphenicol acetyl transferase gene (CAT) for assay by transfection into HTLV-III infected and uninfected cells. The location of *cis* acting regulatory elements could then be deduced from the analysis of deletion clones in LTR which fail to produce a CAT expression. *Trans* acting regulatory elements can be identified by co-transfection of LTR-CAT with other plasmid constructs not linked to CAT but containing deletions in LTR, *tat* or *art*/*trs*.

4. Experimental Methods

Plasmid Clones

Plasmid pSVOCAT, pSV2CAT, and RSVCAT were obtained from Dr. B. Howard (22). Clone pC15CAT was constructed by blunt ending the Pst I cDNA insert of clone pC15 (17) with T4 DNA polymerase and the four NTPs (Fig. 4). Hind III linkers were added and the resulting fragment was cloned into the Hind III site of pSVO. DNA sequence analysis confirmed that the entire 3' region of the C15cDNA insert was present in C15CAT including the poly A tail. VHHCAT was constructed by cloning the most 3' Hind III fragment of the HXB2 provirus (24) into the Hind III site of PSVO.

Synthetic Oligonucleotides

Oligonucleotides were kindly synthesized by L. Lee of Program Resources Inc., Frederick, MD.

Construction of 3' deletion Mutants

The clone -117 (Fig. 4) was cleaved with either Bgl II + Sst I to create -117 BS or Sst I for -117 S. The DNAs

were blunt ended as described above then self ligated and transfected into the bacterial strain HB101 (BRL). Clone -117 +56 was created by ligating an Sst I to Hind III synthetic oligonucleotide fragment containing the sequences +39 to +56 from the CAP site into the Sst I and Hind III sites of clone -117.

CAT Assays

CAT assays were performed as described previously (24).

DNA Sequence analyses

The 5' deletion clones were cleaved with Xba I. The 5' Xba I ends were labeled in a reaction with T4 polynucleotide kinase and ^{32}P-ATP (7000 Ci/mmole, New England Nuclear) and then cleaved with Hind III. The Xba I to Hind III fragments were purified from 6% acrylamide gels and sequenced by the Maxam-Gilbert method. The -65E and -48E clones are cleaved with Bgl II, 5' end labeled as above and then cleaved with Hae II. The Bgl II to Hae II fragments of each were purified as above and then sequenced by the Maxam-Gilbert method (25).

5. Results

LTR-directed expression in infected and uninfected cells

The 5' deletion clones (Fig. 5) were transfected into H9 cells or H9/III cells and CAT assays performed on the cell lysates within 40-48 hours post transfection. The results of the assays are shown for each clone in Fig. 5. The first column shows values for H9 cells and the second column shows the results in H9/III cells. The much increased level of CAT activity in infected cells is apparent from the reaction conditions. The RSVCAT plasmid (values in parenthesis) served as a positive control in these experiments for both transfection efficiency and CAT expression. The values for infected cells show >1000 fold activity as opposed to uninfected cells for clones containing complete LTRs.

Deletions to position -117 showed no significant differences (<2 fold variation) from CD12CAT in either infected or uninfected cells. However, the activity of clone -103 activity in uninfected cells was reduced by a factor of 4-9 fold while values for infected cells did not significantly change. Further deletions to -65 and -48 resulted in loss of activity in uninfected cells and either partial or complete loss of activity in infected cells respectively.

Lack of a Negative Regulatory Element

Negative regulatory element were not detected in the U3 region in this study in contrast to previous studies (19). The reasons for this are unclear. The 6 fold suppression of activation reported would still allow greater than 1000 fold activity in HTLV-III infected cells relative to uninfected cells reflecting in our opinion a minor role for this element even if it existed.

Localization of enhancer elements

The loss of activity in uninfected cells upon deletion of sequences between -103 and -65 suggested that an enhancer element might be located within this region. To test this hypothesis, an oligonucleotide spanning -104 to -80 was synthesized with Xba 1 compatible ends and cloned into the Xba 1 site of -65 and -48. The resulting clones were assayed for functional restoration. The results of the CAT analyses of these clones are shown in Table 3. Clone -65E2 contained the sequences in the 5' to 3' orientation while clone -65E5 contained the reversed orientation. The direct repeats were inserted ahead of -48 in the 5' to 3' orientation in clones -48E9 and -48E14 and in the 3' to 5' orientation in clone -48E8. DNA sequencing confirmed that the clones contained the entire -105 to -80 region except -48E14 which is missing base pair -80.

In all, plasmids -65E2 and -65E5 were positive within or above the range seen for the fully active LTR sequences of CD12 while -65 (data not shown) and the clones -48E8, -48E9 and -48E14 were approximately 10 fold less active. Since the activities seen with these clones did not depend on the orientation of the inserts the direct repeat region has the property of the enhancer.

The tat response region (TAR)

Previous studies have shown that heterologous promoter and enhancer sequences could be placed upstream of the -17 to +80 region of the HTLV-III-LTR and be activated 1000 fold (19). We have made further deletions in this region to more precisely define the elements which are responsible for the transactivation phenomenon. For this we used the tat responsive clone, -117, to construct a deletion from the Bgl II site at +20 to the Sst I site at +38. The clone, -117 BS, lacked tat responsiveness but displayed normal levels of expression in H9 cells as shown in Table 3. Thus sequences intermediate of Bgl II and Sst I were indispensable for tat response. Furthermore a deletion of four base pairs of the Sst 1 site GAGCTC at +34 was also made. This clone,

-117 S, had no tat response when transfected into H9/III cells (Table 3). Identical results were obtained when clones -117 BS or -117 S were transfected into nonlymphoid COS-1 cells.

To further delimit the region necessary for tat response shown here to extend downstream of base pair +38, a 17 base pair oligonucleotide of sequences from +39 to +55 with 5' Sst I and 3' Hind III insertible ends was synthesized and cloned into the Sst I (+38) and Hind III (+79) +55. The resulting clone, -117 +56H was restored for tat response (data not shown). Inspection of the sequences in the region of the Bgl II and Sst I sites for significant features revealed an inverted repeat sequence of 11 and 10 base pairs at positions +12 and +40. An 8 bp directly repeated sequence (CTCTCTGG) was also present at +5 and +37. Simply preserving the inverted repeat sequences as in -117 BS was not sufficient for maintaining tat responsiveness. The larger stem loop structure may be important because of the spacing between the inverted repeat. On the other hand, the second CTCTCTGG sequence was disrupted in both -117 S and -117 BS and it is possible that this sequence may be important for tat response.

6. Discussion

The experiments described here have identified an enhancer element of 26 base pairs within the HIV-LTR. This enhancer is localized at position -105 to -80 from the CAP site. Since the activities seen with the clones in this region did not depend on the orientation of the inserts, the direct repeat region has the property of the enhancer. A perfect homology of the 10 base pair direct repeat was found in the SV40 72 base pair enhancer region in the sense of the late mRNA and in the 18 base pair repeated motif present in the cytomegalovirus major immediate early gene promoter (26). Since the corresponding region of HTLV-III is biologically active, these elements may have a similar biological role in the regulation of cytomegalovirus and SV40 expression as well. Our analyses show that deletion of the first G of the enhancer is not critical since clone -103 is active. Mutational analyses of the homologous sequences in SV40 showed that the last cytosine residue is critical for SV40 enhancer function. It will be of interest to check similar mutations in HTLV-III for activity.

The presence of sequences at -65 were found to be crucial for high level activity along with the enhancer. Sequences within the region are known to bind the nuclear factor Sp-1. The presence of viral encoded potentiators in the H9/III cell system (ie the tat III protein) may provide a powerful means of measuring very subtle changes in low level trans-

criptional activity (see below). Alternatively, tat III activation may include an interaction with Spl or its target sequence. These possibilities are being investigated in cell free transcription systems and DNA binding experiments. We conclude that the major regulatory elements involved in the region upstream from the HTLV-III promoter consist of an enhancer region from -104 to -80 which acts cooperatively with the Spl binding site at -65, site II.

The region downstream of the promoter, the TAR or trans-acting response region is necessary for elevated levels of CAT expression in HIV infected cells relative to that in uninfected cells. The TAR region had been previously mapped to position -17 to +80, but the experiments described here indicate that sequences downstream of +49 are not essential for tat response.

It has been demonstrated recently that a major effect of tat activation occurs post transcriptionally (18). We report that any structural feature in the HTLV-III mRNA which would play a role in tat response would have to be located upstream of +55. It is conceivable that the tat protein or cellular factors induced by the virus could bind to such a structure and facilitate post transcriptional enhancement of expression.

C. MOLECULAR STUDIES OF STLV-III$_{AGM}$ (SIV) AND OTHER RELATED RETROVIRUSES

1. Statement of Problem

Acquired Immune Deficiency Syndrome (AIDS) is caused by infection with the human retrovirus called human T-Cell lymphotropic virus type III (HTLV-III) or human immunodeficiency virus (HIV-1). Epidemiological studies indicate that the virus originated in Africa. The discovery of a retrovirus distantly related to HIV which infects a major population of African Green Monkeys has led to the hypothesis that the simian virus (SIV) may be the ancestor of HIV or that the two viruses might share a common ancestor. Although SIV is apparently non-pathogenic in African Green Monkeys, it preferentially infects and kills OKT4+ lymphocytes *in vitro* in the same manner as HIV-1 and it can induce an AIDS-like disorder in captive macaques. A fourth human retrovirus has been isolated independently from West African patients by several laboratory groups. This virus, designated HIV-2 demonstrated a close relationship to SIV and a distant relationship to HIV-1. This virus is tropic but non-cytopathic for OKT4+ lymphocytes *in vitro* and is reportedly non-pathogenic *in vivo*.

Several distinct proviral clones of HIV-1 have been shown to be infectious and cytopathic. The nucleotide sequence of several of these HIV-1 isolates have been determined. Analysis of the DNA sequence showed the existence of at least eight genes: *gag*, *pol*, *env*, *sor*, *3' orf*, *R*, *art/trs* and *tat-3*. It is imperative that the structure and function of each of these genes and their products be defined. In order to understand the evolution and function of the HIV virus, the molecular structure and functional regions of the new viral isolates STLV-III$_{AGM}$ (SIV) and HIV-2 must be determined. Since SIV and HIV-2 viruses have been shown to be serologically cross-reactive with some HIV-1 proteins, they offer valuable comparative models for HIV-1 pathogenesis with direct implications in the development of therapeutic and vaccine reagents.

2. Background

Since 1981, when the first cases of the acquired immunodeficiency syndrome (AIDS) were reported (27), two new groups of human retroviruses have been discovered and characterized. The first, HIV-1 (28, 29, 30), is prevalent in AIDS cases worldwide and seem to be the primary cause of the severe T-cell depletion observed in patients with AIDS. The existence of the second group of human retroviruses was first shown by a serological study of a healthy West African population using as target antigens the proteins of a retrovirus dis-

covered in African Green Monkeys ($STLVIII_{AGM}$) by the serological cross-reactivity of its major core protein with that of HIV-1 (31). Several viral isolates have been obtained from the peripheral blood of infected people and designated HTLV-IV (32). Simultaneously, a group of West African patients with frank AIDS, were found to be infected with a virus related to STLV-III (33,34) but not to HIV-1. A third group (35) reported the isolation from a West African patient of a new retrovirus which was closely related to $STLV\text{-}III_{AGM}$ but distantly related to HIV-1. This virus was designated as SBL6669 and is a member of the HIV-2 group of human retroviruses.

All of these new viruses are immunologically related; whereas, they are only distantly related to HIV-1. Like HIV-1 (36, 37), they appear to use the CD4 molecule as a receptor on the cellular membrane (31,32,33). Direct experimental evidence, though, has been provided only for $STLV\text{-}III_{AGM}$ (38). We have performed molecular analysis of several of these human and non-human primate retroviruses and have found them to be closely related. $SLTV\text{-}III_{AGM}$ is essentially identical to some of the HTLV-IV isolates (38,39,40); whereas, extensive polymorphism has been demonstrated among other HTLV-IV (B. Hahn and G. Shaw, personal communication) and HIV-2 isolates (34,41). The inferred amino acid sequence of the proteins of $STLV\text{-}III_{AGM}$ (41,42) exhibits an overall 40% amino acid homology with HIV-1. The HIV-2 isolate (43) is as distant from HIV-1 (42%) by amino acid sequence homology as $STLV\text{-}III_{AGM}$. Conversely, $HIV\text{-}2_{rod}$ shares a much closer (72%) amino acid homology with $STLV\text{-}III_{AGM}$ (44) indicating that $STLV\text{-}III_{AGM}$ is a member of the HIV-2 group.

3. Rationale

What is not currently clear is the degree and manner of divergence among various HIV-2 isolates. This is a critical issue since the pathogenic potential of these viruses is currently being debated. In order to investigate the relationship among different HIV-2 isolates we have molecularly cloned SIV and a member of the HIV-2 group. The nucleic acid sequence of SIV and $HIV\text{-}2_{NIH\text{-}Z}$ has been obtained and compared with $HIV\text{-}2_{rod}$, HIV-1, and the ungulate retroviruses Visna and EIAV. A comparison of the nucleotide and amino acid sequences of the new viral isolates with each other, HIV-1 and other retroviruses should help to determine the evolution of HIV and will aide in the investigation of the biological properties of this group of retroviruses.

4. Experimental Methods

Cloning of Proviral DNA

High molecular weight DNA was obtained from the HIV-2_{NIH-Z} infected cell line HUT 78 and partially cleaved with BamHl. Different fragments of DNA were size selected on a sucrose gradient and the DNA fraction containing 20 Kb fragments was purified by ethanol precipitation. The DNA was ligated into the arms of EMBL-3 vector, previously cleaved with BamHl and the ligated DNA was packaged *in vitro*. The recombinant phages were plated and screened using as a probe the SS35 and B16 plasmids containing the *env* and *gag* sequences of STLV-III_{AGM}, respectively (41).

The proviral DNA of STLV-III_{AGM} was obtained from a genomic library constructed from the DNA of the infected cell line K6W using the lambda phage vector EMBL-3. Nine clones containing various overlapping portions of the STLV-III_{AGM} genome were purified and characterized by restriction endonuclease analysis. Five of them were used to generate viral DNA fragments for subcloning in appropriate vectors. Standard DNA cloning methods were used as in Maniatis et al., 1981 (45).

DNA Sequencing

DNA restriction fragments of the provirus were removed from the genomic phage clones and were subcloned into the appropriate enzyme sites of the plasmid vector Bluescribe. Other subclones were generated by inserting DNA restriction enzyme fragments of HIV-2_{NIH-Z} and STLV-III_{AGM} into M13 bacteriophages, mp8 and mp9. The DNA sequence was obtained by the dideoxy chain termination procedure (46) using synthetic primers and both the Klenow fragment of *E. coli* polymerase and the T7 DNA polymerase on single and double stranded DNA. Approximately 3 KB of the HIV-2_{NIH-Z} proviral DNA was also sequenced by the chemical degradation method of Maxam and Gilbert (25). The nucleotide sequence was organized with the help of the computer program Microgen (47).

5. Results

a. Genetic Analysis of STLV-III_{AGM} (SIV)

The endonuclease restriction map of the STLV-III_{AGM} provirus derived from overlapping phage clones is depicted in the lower part of Figure 6. The DNA sequence of the entire provirus was obtained by the dideoxy chain termination method and primer extension using both single stranded and double stranded template DNA (46). The complete sequence of the STLV-III_{AGM} provirus, derived by combining the DNA sequence of

the lambda phage clones, is presented in Figure 2 and is numbered from the 5' boundary of the R region of the upstream LTR to the 3' end of the R region of the downstream LTR. The overall homology at the nucleotide sequence level between STLV-III$_{AGM}$ and HIV-1 is 55%.

Long Terminal Repeat (LTR)

The STLV-III$_{AGM}$ LTR is 800 base pairs long. To define the boundaries of the LTRs, we compared the junction sequence of the U3-cellular sequence of the 5' LTR and the U3-3'- orf sequence at the 3' end of the virus. Similarly, the end of the U5 was identified by comparing the sequence of the 3' viral-cellular junction and the 5' LTR-gag junction. The boundaries of the R region with respect to the U3 and U5 were determined by comparison with the HIV-1 LTRs, (48) and the sequence of several STLV-III$_{AGM}$ cDNA clones (Colombini-Hatch et al., personal communication). The sizes of the U3, R and U5 are 498, 176 and 126 base pairs, respectively. The sequence TATAA, found 25 nucleotides upstream from the start of transcription in many eukaryotic genes, is located in the U3 at position -25 from the presumed boundary of the U3 and R regions. The consenus sequence, AATAAA, which signals the addition of the polyadenylate tail, is located at position 152 to position 157.

The primer binding site (PBS), located downstream from the U5 of the 5'LTR, is complementary to the same isoaccepting form of transfer RNA (tRNA$_{lys3}$) as the PBS of HTLV-IIIB and other HIV-1 strains (48). Surprisingly, in view of the close relatedness of these viruses, the STLV-III$_{AGM}$ LTR is much longer than the HIV-1 LTR (800 versus 634 base pairs). The STLV-III$_{AGM}$ and HIV-2$_{rod}$ LTRs are closer in size to the HTLV-I (49) and HTLV-II LTRs, and are much larger than the LTRs of the ungulate lentiviruses. The greater size of the STLV-III$_{AGM}$ LTR as well as the HIV-2 LTR relative to that of HIV-1 is due to the presence of several regions in the STLV-III$_{AGM}$ of sequences not found in the HIV-1 LTR.

The gag and pol genes

Retroviral core proteins are derived from the proteolytic cleavage of a polypeptide precursor encoded by an open reading frame at the 5' end of the genome. The first large open reading frame in the STLV-III$_{AGM}$ genome starts at nucleotide 539 (Figure 7) and could encode 506 amino acids. The length of the gag precursor polypeptide reported for different HIV-1 strains varies between 500 and 512 amino acids. Alignment of STLV-III$_{AGM}$ and HTLV-IIIB amino acid sequences suggests that its p17 homology contains 135 amino acids and is likely generated by cleavage between Tyr and Pro at position 941-946.

Proline is the first amino acid at the N-terminus of HIV-1 p24, the p26 of the ungulate lentiviruses EIAV and Visna, the HTLV-I p24, and the HIV-2_{rod} p26, and is therefore very likely to occur at the same position in the STLV-III_{AGM} p24. The overall homology of the gag precursors of STLV-III_{AGM} to those of HTLV-IIIB and HIV-2_{rod} is 51% and 82% respectively (see Table 4). The highest amino acid identity can be found between the major core proteins of STLV-III_{AGM} and HTLV-IIIB (66%) and HIV-2_{rod} (88%). These data are consistent with the strong cross reactivity of these proteins in radioimmune assays and Western Blots. A comparison of the amino acid sequences of the major core protein of the lentiviruses equine infectious anemia virus (EIAV) and visna with the STLV-III_{AGM} and HIV-1 gag proteins shows a significantly higher homology of EIAV to STLV-III_{AGM} (29%) and HIV-1 (29%), than of visna to STLV-III_{AGM} (20%) or HIV-1 (19%).

At the carboxy terminus of the HTLV-IIIB gag precursor, two repeated sequences have been described. The first, an imperfect repeat encoding 12 amino acids, is also present in the STLV-III_{AGM} genome (Figure 7), as well as in the HIV-2_{rod} genome at positions 1605-1641 and 1667-1703, suggesting that this sequence duplication must have occurred long ago in a common ancestor of these three viruses. The second, present in HTLV-IIIB, is a perfect repeat that also encodes 12 amino acids (4), and is absent in the STLV-III_{AGM} genome. The fact that this repeat in HTLV-IIIB is perfect and that some other HIV-1 isolates as well as STLV-III_{AGM} lack it, strongly suggests that the duplication in the HTLV-IIIB genome must have occurred relatively recently. Interestingly, an imperfect direct repeat of 18 amino acids can be detected in the HIV-2_{rod} genome, indicating that the border between the gag and the pol open reading frame may be particularly prone to duplications. We have observed similar duplications in the genome of other HIV-1 isolates (unpublished results).

The second large open reading frame in the STLV-III_{AGM} genome spans from nucleotides 1714 to 4875, has a coding potential of 1053 amino acids, and overlaps the gag gene by 342 bases (Figure 7). We assume that either a splicing or a frameshift event in this area of overlaps results in the translation of the pol gene. By analogy with HIV-1 and other retroviruses, this open reading frame should contain the genetic information for protease, reverse transcriptase and endonuclease proteins. In HTLV-IIIB, the pol region encodes 1003 amino acids and also extensively overlaps the gag reading frame. The amino terminus of the HTLV-IIIB reverse transcriptase, beginning with a Pro-Ile, has been identified by amino acid sequencing of the purified p66/p51 viral enzyme (50). A perfect alignment can be found between the amino termini of the HTLV-IIIB and HIV-2_{rod} reverse transcriptase and

the inferred amino acid sequence of the STLV-III$_{AGM}$ RT gene. Thus, it is likely that the Pro-Ile in position 2326-2331 represents the first two amino acids of the STLV-III$_{AGM}$ reverse transcriptase. The overall amino acid homology of the reverse transcriptase of STLV-III$_{AGM}$ versus HTLV-IIIB and HIV-2$_{rod}$ is 53% and 76% respectively (see Table 4). The amino acid sequence upstream from the beginning of the reverse transcriptase is also well conserved between HTLV-IIIB and STLV-III$_{AGM}$ and is probably the region which encodes for the protease protein. A comparison of the amino acid sequences of that region among HTLV-IIIB, EIAV and visna virus leads to the conclusion that in these viruses the protease is encoded within the same open reading frame that codes for the reverse transcriptase gene. The high degree of relatedness of the STLV-III$_{AGM}$ amino acid sequence in this region with that of HIV-2$_{rod}$ (88%), HTLV-IIIB (49%), EIAV (37%), and visna virus (30%) allows a more precise localization of the presumed cleavage site (Leu-Pro) between protease and reverse transcriptase genes.

Open Reading Frames in the Middle of the Genome

Three open reading frames have been identified in the middle portion of the STLV-III$_{AGM}$ genome. The first sor, (also referred to as Q), overlaps by 97 nucleotides with the end of the pol gene and could encode for a 225 amino acid protein. A potential methionine initiation codon is present 12 amino acid downstream from the beginning. This protein shares 24% amino acid homology with the sor gene of HTLV-IIIB and other HIV-1 strains. Since sor appears to be very well conserved among the various HIV-1 isolates and since STLV-III$_{AGM}$ and all HIV-1 isolates so far analyzed are, in general, closely related, the percentage of homology of the STLV-III$_{AGM}$ sor protein with that of HTLV-IIIB is surprisingly low. However, the amino acid hydropathy profile of the simian and human viral proteins is very similar (data not shown). A higher homology of 64% can be found between the sor (Q) of HIV-2$_{rod}$ and STLV-III.

The second short open reading frame overlaps with sor by 171 bases and with the amino terminus of the env gene by 21 bases and spans from position 5277 and 5723. It could encode 149 amino acids and has a possible intiation codon at position 2 (Figure 8). The first 60 amino acids share high homology with the X open reading frame described in the HIV-2$_{rod}$ genome (43). The last part of the sequence shares considerable homology (40%) with the second exon of the transactivator gene (tat-3) of HTLV-III which includes the functionally important domains of the tat-3 protein. The amino acids conserved between the STLV-III$_{AGM}$ tat-3 and the HTLV-IIIB tat-3 are indicated by asterisks in the lower part of Figure 8. A comparison of this portion of the STLV-III$_{AGM}$

genome with the published HIV-2_{rod} sequence indicated that the STLV-III_{AGM} genome obtained from the K6W cell line is lacking 350 nucleotides in the middle region. The deletion, as it is schematically represented in Figure 8 includes part of the X region, most of the R gene and the first 14 amino acids at the amino terminus of the tat-3 protein. Hybridization of an oligomeric probe for this deleted region to the cellular DNA from which these clones were derived indicated that these sequences were also absent in the majority of the proviruses in the infected cells. This suggests that STLV-III may have a propensity for deletion of regions from the central area of the genome. Since we have not shown these clones to be biologically active, we cannot say whether or not such deletions would result in the loss of biological activity.

The third open reading frame (5774-5987) encodes for 71 amino acids and has a potential initiating methionine in position 3 (Figure 8). This open reading frame overlaps completely with the coding region for the amino terminus of the large envelope protein gp120.

Conserved domains in the envelope of STLV-III_{AGM}, HIV-1 and HIV-2

The largest open reading frame in the 3' region of the STLV-III_{AGM} genome corresponds to the gene encoding the envelope protein. The first residue is at position 5702 and the probable initiating methionine codon is located 10 amino acids downstream. The env gene probably extends to a termination codon at base 8372-8374, and could encode a peptide of 881 amino acids. There is also a termination codon at base 7934. Based on protein alignments with the HTLV-IIIB env gene, this termination codon is located in the transmembrane portion of the envelope gene immediately prior to the splice acceptor site in the third exon of tat-3 mRNA. Interestingly, the homology between STLV-III_{AGM} and HTLV-IIIB decreases considerably after the termination codon at position 7934 of the STLV-III_{AGM} genome, implying that these sequences may not be important for viral replication. We have observed the termination codon at 7934 in several independent STLV-III_{AGM} clones and the same termination codon is present in a biologically active HTLV-IV clone (38). It is possible that its presence has biological significance in the pathogenesis of STLV-III infection in vivo. Furthermore, although the presence of a termination codon is not reported in the published sequence of HIV-2_{rod}, it appears to be present also in some HIV-2_{rod} clones in the same position (Alizon, M., personal communication). Comparative analysis of the envelope proteins of STLV-III and HIV-2_{rod} indicated that 70% of the amino acids are identical in both the extracellular and the transmembrane portion of the protein. Higher amino acid

homology can be found between the cleavage site and the termination codon (82%) than at the carboxy terminus of the envelope open reading frame (54%). It has been previously shown that deletion of part of the carboxy terminus leads to a reduction of the cytopathic effect in vitro. The presence of this termination codon in STLV-III_{AGM}, HTLV-IV and HIV-2, which would eliminate 40% of the transmembrane protein, could modulate the pathogenicity of the virus through variability in the degree of suppression of termination in vivo.

An alignment of the inferred amino acid sequences of the STLV-III_{AGM} and HTLV-IIIB env gene shows 38% homology overall (Figure 9), and permits us to predict that the signal peptide extends to the serine at base 5810-5812 and that the large and small envelope proteins are cleaved, as in HTLV-IIIB, between the arginine and glycine residues at bases 7307-7312. Based on this assumption, the homologies of the two large env proteins and transmembrane are 35% and 44% respectively. The hydropathy profile for the putative STLV-III_{AGM} transmembrane protein is almost identical to that of HTLV-IIIB, consistent with functional similarity. Much higher homology can be found between the envelope proteins of STLV-III_{AGM} and HIV-2_{rod} (Table 5).

There are several remarkable aspects of the homology of the STLV-III_{AGM} and HIV-1 env proteins which may reflect on the relationship of their structure and common function (s). First, there is an extremely strong conservation of cysteine residues. Of the twenty-two cysteines present in the HTLV-IIIB env protein (Figure 9) 21 are present in the same position in the STLV-III_{AGM} env protein. Different strains of HIV-1, which vary by up to 20-25% among themselves in the amino acid sequences of the env protein, similarly show an invariant conservation of these cysteines. This suggests that disulfide bridges play a critical role in maintaining the three-dimensional structure of these env proteins. Similarly, 26 out of 27 cysteines are conserved between the envelope proteins of STLV-III_{AGM} and HIV-2_{ROD} (Figure 9).

The most interesting aspect of the localized homologies between the env genes of STLV-III, HIV-1, and HIV-2_{rod} is that they occur in those regions which are most conserved among different strains of HIV-1. Conversely, the env genes of HTLV-IIIB and STLV-III_{AGM} are most divergent in the regions of the HIV-1 env which are most hypervariable. This suggests that the conserved regions are important for functions which are common to the functions of the env proteins of these retroviruses. Since STLV-III_{AGM} like HIV-1 and HIV-2_{rod}, seem to use the CD4 protein as their receptor, this protein may bind to highly conserved regions. In fact, it has recently been shown by Lasky et al. (personal communication) that

the region containing the last cysteine in HIV-1 gp120 is critical for binding to the CD4 molecule. This region spans from amino acid 442 to 463 in the STLV-III_{AGM} envelope and has an amino acid sequence almost identical in HTLV-IIIB, STLV-III_{AGM}, and HIV-2_{rod} (Boxed region in Figure 9). Another important antigenic site has been mapped in this highly conserved region. Cease et al. (51) have identified two peptides, T1 and T2, which elicit T cell immunity. The 16 amino acids T1 peptide maps within the putative envelope region binding site of the CD4 molecule.

The open reading frame homologous to the Trs gene

In the region corresponding to the transmembrane portion of the putative STLV-III env gene an open reading frame encoding 96 amino acids can be found between nucleotides 7900 and 8187. This region follows immediately after a splice junction present in a functional STLV-III_{AGM} tat-3 cDNA (Colombini-Hatch et al., personal communication). The probable protein encoded by this region shares 29% homology with an equivalent open reading frame in the HTLV-IIIB genome. In HIV-1, the existence of this gene was discovered by in vitro mutagenesis of a biologically active HIV-1 clone and designed art or trs (52,53). The protein product of the HIV-1 trs of 19 Kd, appears to be essential for the synthesis of gp120 (Knight et al., 1987).

The 3'orf Gene

The last large open reading frame is located at the 3' region of the viral genome and overlaps with the carboxy terminus of the envelope gene and the U3 region of the viral LTR. This region (8196-8828) has a coding potential for 211 amino acids and a potential initiating AUG codon at position 8208. A similar region has been described in the HTLV-IIIB genome which has been shown to encode a 27K protein (54,55). The overall amino acid homology between this putative STLV-III_{AGM} protein and the HTLV-IIIB 3'orf is 34%. A more detailed analysis of the amino acid composition of the 3' orf proteins among the STLV-III_{AGM} and the American, French and African HIV-1 isolates and HIV-2_{rod} reveals the existence of a highly conserved region of 70% homology spanning about 80 amino acids (Figure 10). The hydropathy profile of this region in the different viruses is almost indistinguishable. It is likely that these 80 amino acids represent the core of the functional domain of the 3' orf protein.

b. Genetic Analysis of HIV-2_{NIH-Z}

The HIV-2_{NIH-Z} virus isolate was obtained by D. Zaguri from a patient with immunodeficiency who originally lived in

Guinea Bissau. A genomic library was constructed from the DNA of the HIV-2_{NIH-Z} infected human neoplastic T-cell line HUT 78 (56). Using the STLV-III_{AGM} proviral DNA as a probe (39), we purified a molecular clone of the complete provirus and determined its nucleotide sequence. The genome of HIV-2_{NIH-Z} is 9431 base pairs long and the overall organization of the open reading frames is consistent with the order 5' LTR-gag-pol-middle region-env-3'orf-3'LTR. Similar structures have been reported for HIV-1 (4,57,58,3). A single structural difference lies in the presence in HIV-2_{NIH-Z} of an extra open reading frame in the middle of the genome which is absent in the HIV-1 genome but which is also present in HIV-2_{rod} and STLV-III_{AGM}, called X (43,41).

The LTR

The HIV-2_{NIH-Z} is 632 nucleotides long. The size of the individual components was derived by analogy with HIV-2_{rod} and STLV-III_{AGM}. The U3, R, and U5 are 329, 176 and 127 base pairs in size, respectively. While the U5 and R are comparable in size to the HIV-2_{rod} and STLV-III_{AGM}, the U3 of the HIV-2_{NIH-Z} is shorter than that of either virus. A sequence alignment of the U3 region of the HIV-2_{NIH-Z} showed a deletion of 228 nucleotides in the HIV-2_{NIH-Z} U3 region. The deletion occurred 60 nucleotides 3' of the polypurine tract (Figure 11). Since the deletion is present in both the 5' and 3' LTRs, it is probably present in the provirus and is not a cloning artifact. The position of the regulatory sequences in the HIV-2_{NIH-Z} LTR, such as the TATAA box, the polyadenylation signal AATAAA, and the SP1 binding sites, are indicated in Figure 11.

The gag and pol gene proteins: relationship to other lentiviruses

The first open reading frame derived from the nucleotide sequence of the HIV-2_{NIH-Z} corresponds to the gag precursor. In the infected cells the size of the gag precursor appears to be 55 Kd as judged by immunoprecipitation with human sera from infected individuals (35). The gag open reading frame has a coding capacity of 519 amino acids. Although definitive evidence has not been provided yet, the gag precursor should exhibit the same size as in HIV-1, p55 (50). The amino acid homologies of HIV-2_{NIH-Z} gag precursor polypeptides when compared with HIV-2_{rod}, STLV-III_{AGM} and the HTLV-IIIB strain of HIV-1 are 92%, 82%, and 52% respectively (Table 5). This indicates that the two human HIV-2s are somewhat more closely related to each other than to STLV-III_{AGM} and that HIV-1 is much more distantly related to all of them. The cleavage site for the major core protein (p24) in the HIV-2_{NIH-Z} gag precursor can be provisionally assigned by alignment of the amino acid

sequences of the gag precursor polypeptides of HIV-2, HIV-1 and STLV-III$_{AGM}$ (Figure 12).

These alignments were performed using the algorithm of Dayhoff and colleagues. These sequences were aligned with those of the major gag proteins of Visna and EIAV in order to establish their respective phylogenetic relationship (Figure 12). The results expressed as percentage of amino acid sequence are summarized in Table 6. The analysis of the p24-26 proteins indicates that HIV-2$_{NIH-Z}$ and HIV-2$_{rod}$ are highly related (96%) and are equally distant from STLV-III$_{AGM}$ (88%). Both HIV-2s and STLV-III$_{AGM}$ appear to be equally distant from HIV-1 (66-68%). Equal homology (28-29%) is found between EIAV and HIV-1, both HIV-2 isolates, and STLV-III$_{AGM}$. Similarly, Visna virus shares only 24-26% homology with HIV-1, HIV-2 and STLV-III$_{AGM}$. Thus, the comparison suggests an ancestral relationship between the ungulate retroviruses and the primate immunodeficiency viruses, but also indicates that the divergence of the primate immunodeficiency viruses occurred much later while HIV-2 and STLV-III$_{AGM}$ diverged later still.

The pol open reading frame of 1190 amino acids (Figure 11) overlaps the gag precursor open reading frame as in HIV-1. The overall amino acid sequence relationships among the pol genes of both HIV-2 isolates, HIV-1 and STLV-III are like those of the gag genes and lead to the same conclusions.

Conserved and variable domains in the envelope of HIV-1, HIV-2 and STLV-III$_{AGM}$

The third major open reading frame in the HIV-2$_{NIH-Z}$ provirus corresponds to the envelope protein (Figure 12) and potentially encodes for 856 amino acids. A comparative analysis of the envelope protein of HIV-2$_{NIH-Z}$ with HIV-2$_{rod}$ and STLV-III$_{AGM}$ showed an overall homology of 80% and 70% respectively. The degree of conservation is comparable in the extracellular and transmembrane envelope proteins (Table 5). The envelope proteins of HIV-2$_{NIH-Z}$ and HIV-1 are much less related (35%), as shown in Table 2. A homology comparison between HIV-1 and HIV-2 was undertaken in order to identify conserved regions which may be crucial for the function of the envelope protein. The position of the cysteines is highly conserved among all these retroviruses. In the extracellular envelope protein 22 cysteine residues are conserved among both HIV-2 isolates and STLV-III$_{AGM}$ and 19 of 22 are also conserved in the same position in all strains reported of HIV-1 (Figure 13).

Similarly in the transmembrane portion of the envelope protein three cysteines are conserved among both strains of

HIV-2 and STLV-III$_{AGM}$ and two are also conserved in the HTLV-IIIB strain of HIV-1. Clearly, disulfide bonds must play a crucial role in maintaining the secondary structure of the envelope proteins.

Further analyses of the amino acids homology among the envelope proteins, identified regions in which either complete amino acid identity or only conservative changes could be detected in all these viruses. Of these envelope conserved regions (ECR) indicated in Figure 13, the ECR-6 which is located in the extracellular glycoprotein has been identified as a putative binding site to the CD4 molecule which is an essential part of the cellular receptor for HIV-1, STLV-III$_{AGM}$ (36,37,38) and most likely HIV-2. A peptide (T-1) capable of inducing cell mediated immunity in mice has been synthesized using the sequence of the ECR-6 region from HIV-1 (51). Regions ECR-7 and ECR-8 which are located at the amino terminus of the transmembrane envelope protein may represent regions involved in the repetitive folding of the protein within the cellular membrane, as proposed by computer assisted analyses of the envelope proteins of different strains of HIV-1 (59). Based on the homology of the ECR7 region with the fusion peptide of human paramyxoviruses, measles, and respiratory syncytial viruses (60), it has been hypothesized that the first 11 amino acids conserved in ECR7 may be the fusion peptide of HIV-2, HIV-1 and STLV-III$_{AGM}$. Finally, the 3' most conserved region in the transmembrane envelope protein (ECR-12) has been implicated to be involved in the cytopathic effect *in vitro* (61) even though others (52) have reported contrary results. Despite these differences, which may be related to the use of different target cells, the conservation of these 17 amino acids in HIV-1, HIV-2 and STLV-III$_{AGM}$ may be biologically significant.

In HIV-2$_{NIH-Z}$ infected cells a protein of 33 Kd (compared to 41 Kd in HIV-1 infected cells) believed to be the transmembrane envelope protein has been identified (unpublished data). Similarly, a truncated form of the transmembrane protein has been identified in STLV-III$_{AGM}$ infected cells (31). The *env* gene of STLV-III$_{AGM}$ contained a termination codon that would eliminate the last 146 amino acids at the carboxy terminus of the transmembrane portion of the envelope (41,42). HIV-2$_{NIH-Z}$ does not have a termination codon at the same position. Since the infected cells from which it was isolated appear to express a truncated gp33, the provirus obtained in our laboratory is not representative of the majority of the provirus present. Although the biological significance of the stop codon in some of these viruses is not clear, it might be important in elucidating their biological properties. This stop codon is present in the same position in some clones of HIV-2$_{rod}$ (62). Furthermore, a comparison of

the amino acid sequence of the transmembrane envelope protein of these viruses, shows a decrease in homology after the stop codon (Figure 13 and 14). These data suggest that this stop codon does have biological significance.

The same regions that have been found to vary most in HIV-1 (63) are generally the most variable in both HIV-2 isolates and STLV-III_{AGM} (Figure 13). The first region of variability among the envelope proteins of HIV-2_{NIH-Z}, HIV-2_{rod} and STLV-III_{AGM} spans from amino acid 112 to 190 in HIV-2_{NIH-2} and corresponds to the hypervariable region from aa130 to aa210 in HIV-1 (63,64). The degree of variability in this region of the West African retroviruses, STLV-III_{AGM}, and HIV-1 isolates were analyzed. The percentage of amino acid identity ranged from 30-60%. This equivalence suggests that both groups of viruses may have spread to their present ranges from a limited focus of infection at approximately the same time.

Identification of putative functional domains of other viral proteins.

Several other genes have been identified in the HIV-1 genome by immunological (54,55,65,66) or functional studies (6,52,53,67). The corresponding genes can be identified in the HIV-2 genome and the comparative analyses of their amino acid sequence indicate strongly conserved domains within some of them.

The HIV-2 trs/art gene, which was discovered in HIV-1 by mutagenesis of a biologically active HIV-1 clone (52,53), seems to be crucial for the expression of the HIV-1 envelope protein. Protein sequence alignments of the HIV-2 and HIV-1 trs/art gene products show an arginine rich region in the second coding exon that is conserved among HIV-1, HIV-2 and STLV-III_{AGM} (Figure 15). Similarly, arginine and cysteine rich regions (Figure 15) can be identified in the first coding exon of the tat proteins, which are responsible for the transactivation of virus expression in these viral isolates (40,68). No recognizable conserved regions were detected within the sor (Q) gene although the sor proteins of these viruses share a similar hydropathy profile (not shown). A highly conserved region could, however, be identified in the 3'orf (F) gene although a correct protein alignment of the 3'orf protein product can not be obtained because of the presence of the 228 nucleotide deletion in the U3.

6. Conclusion

The identification of an increasing number of human and simian retroviruses in the last five years, make it imperative to determine their precise genetic relationships in order to

elucidate the genetic basis for their pathogenic effects. Studies on the replicative functions of HIV-1 and its role in T-cell killing have shown that these human retroviruses have a more complicated mechanism of regulation than the non-primate retroviruses. The discovery of a second group of viruses in both primate and human which are structurally and genetically related to HIV-1 and are also associated with immunodeficiency calls for a reinterpretation of the natural history of these viruses and for a reevaluation of the hypothesis that AIDS is a new disease. In fact, analyses of the rate of nucleotide changes suggests that HIV-1 and HIV-2 might have diverged from each other as recently as 40 years ago (G. Meyers, Los Alamos; personal communication) while the first documented cases of AIDS or aggressive form of Kaposi in young people date as far back as the early 1960s (69).

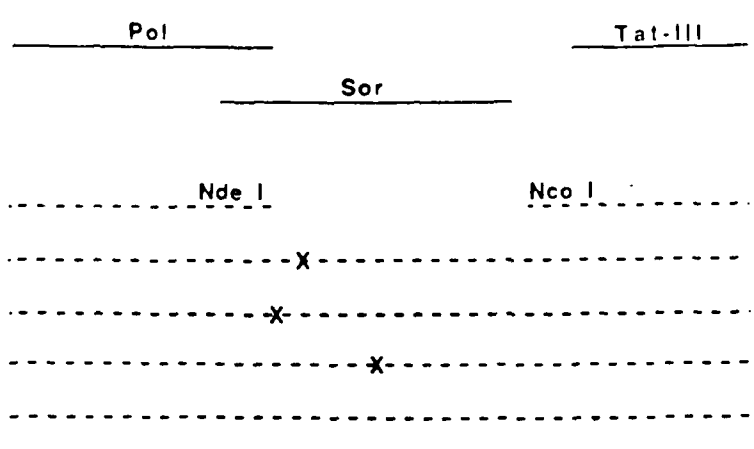
The molecular characteristics of STLV-III and HIV-2_{NIH-Z} reveal that their genetic organization is very similar to HIV-1. STLV-III_{AGM} and HIV-2 are more closely related to each other than to HIV-1. Nevertheless, the extent of similarity among these primate viruses indicate that they arose from a common ancestor. The immune cross reactivity of STLV-III_{AGM}, HIV-1 and HIV-2_{rod} in the major core proteins reflects the high conservation of the gag gene sequence. Comparison of the gag protein of STLV-III_{AGM}, HIV-1 and HIV-2 with other members of the lentiviridae family demonstrate that the former represents a group of viruses that must have diverged from each other more recently than they diverged from the latter group and that STLV-III_{AGM} and HIV-2 diverged from each other more recently than either diverged from HIV-1.

The sor gene of STLV-III_{AGM} as well as of HIV-2 appears to be less conserved than all the other genes. The biological significance of this observation is not clear since the function of the 23 Kd HIV-1 sor protein (70) is unknown.

Conversely, the high degree of conservation of the central region of the 3'orf protein between these three viruses suggests that this region may be the functional domain of the protein. The 27 Kd HIV-1 3'orf protein is a cytoplasmic protein (55) apparently not required for viral replication in vitro (61). The conservation of the amino acid sequence of the 3'orf proteins would suggest its importance in the biology of these primate retroviruses, perhaps as a negative regulatory element.

5 One point of considerable debate is whether the West African viruses (HIV-2) cause AIDS in people. Early reports, which identified a human virus related to STLV-III_{AGM}, on the basis of serology, suggested a lack of disease association (32) while others reported the isolation of HIV-2 viruses from a few patients with immunodeficiency and no signs of infection with HIV-1 (33). More recently, an increasing number of HIV-2 isolates have been obtained from patients with AIDS from West Africa (68, 35). However, a retrospective seroepidemiological study on 4248 people in West Africa showed the absence of any clinical signs in 330 infected people (71). The changes in the envelope proteins of two HIV-2 isolates are similar which suggests that HIV-1 and HIV-2 have existed in their present population for similar lengths of time. Therefore, it is possible that the discrepancy between these studies is due to a lower morbidity rate for HIV-2. A more difficult question is whether there is a fundamental genetic difference between the two virus groups that could explain their apparent different biological behavior in infected individuals. The overall genetic structure of HIV-1 and HIV-2 is very similar, with the exception of an extra open reading frame in HIV-2, which has been designated X. No evidence is currently available which would indicate whether or not X is translated.

The presence of a stop codon in the transmembrane portion of the env gene of some of these viruses opens the possibility that two different forms of the envelope small protein could be synthesized. The biological significance of this phenomena in vitro and in vivo remains to be investigated.



Mutants derived from pHXB2-gpt	Stable infections established after transfection of H9 cells
ΔS	0/6
6.9	0/6
3.3	0/7
153.0	0/3
X	4/4

Fig. 1 Construction and properties of sor mutants of HIV-1. Plasmid X was generated from pHXB2gpt by removal of an EcoRI site in the polylinker region. Mutant S was prepared by digestion with NdeI and NcoI and religation. Mutants 6.9, 3.3, and 153.0 were generated by site directed mutagensis such that stop codons were introduced in frame at the locations marked ----X---- (see text). Mutated fragments were recloned into clone X and transfected into H9 cells by protoplast fusion. Cultures were monitored for HIV 1 production throughout the course of the experiment.

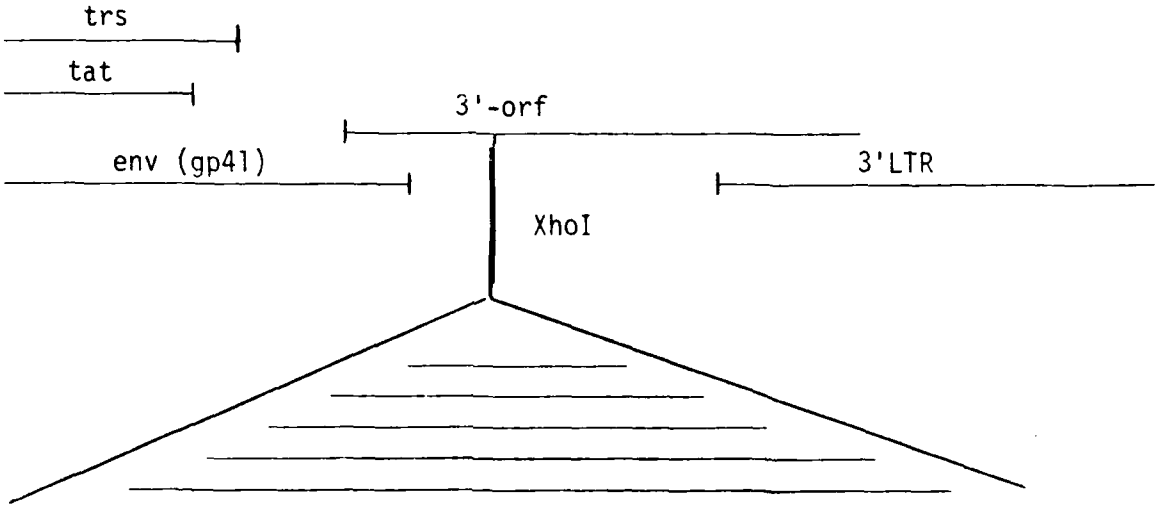


Fig. 2 Construction of HTLV-III clones with deletions in the 3' region of the virus. Plasmid pHXB2gpt was digested with XhoI, treated with Bal31, blunt-ended with T4 DNA polymerase and religated. In some cases a linker arm was added to the blunt ended DNA before redigestion and ligation. Clones in which deletion extended into the gp41 region were selected for further study.

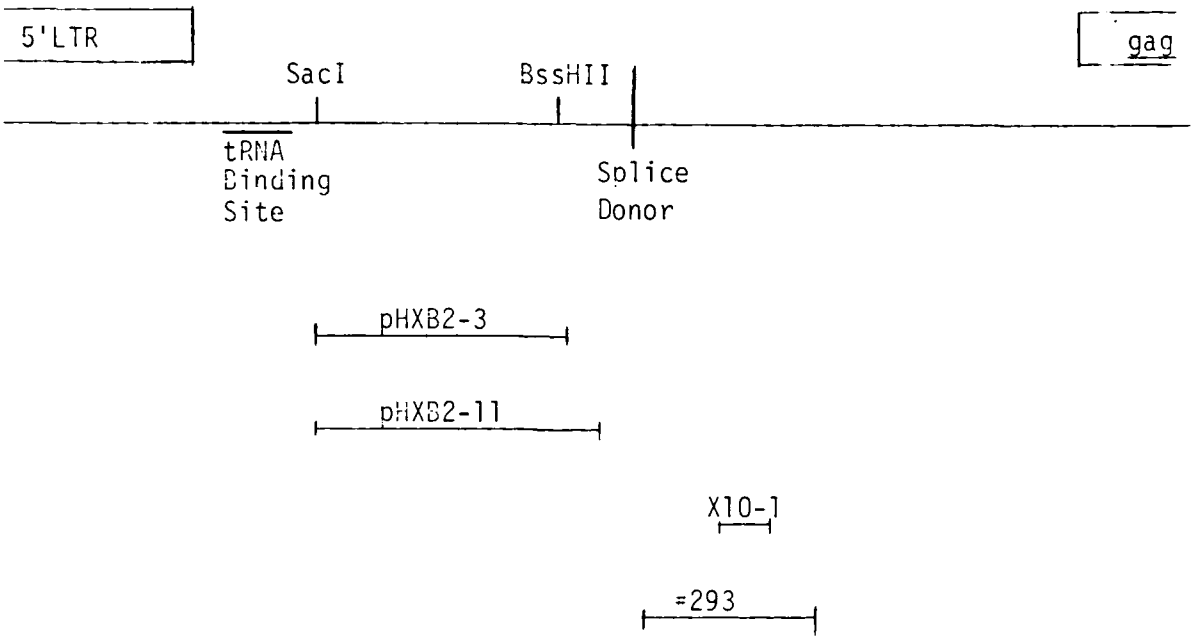


Fig. 3 Construction of packaging defective mutants. A series of deletion mutations in the region between the 5' LTR and gag were constructed from pHXB2gpt. The clone was cut with either SacI or BssH2, subjected to Bal31, blunt ended with DNA polymerase, and religated. Mutants X10-1-25 and #293 were obtained by site directed mutagenesis.

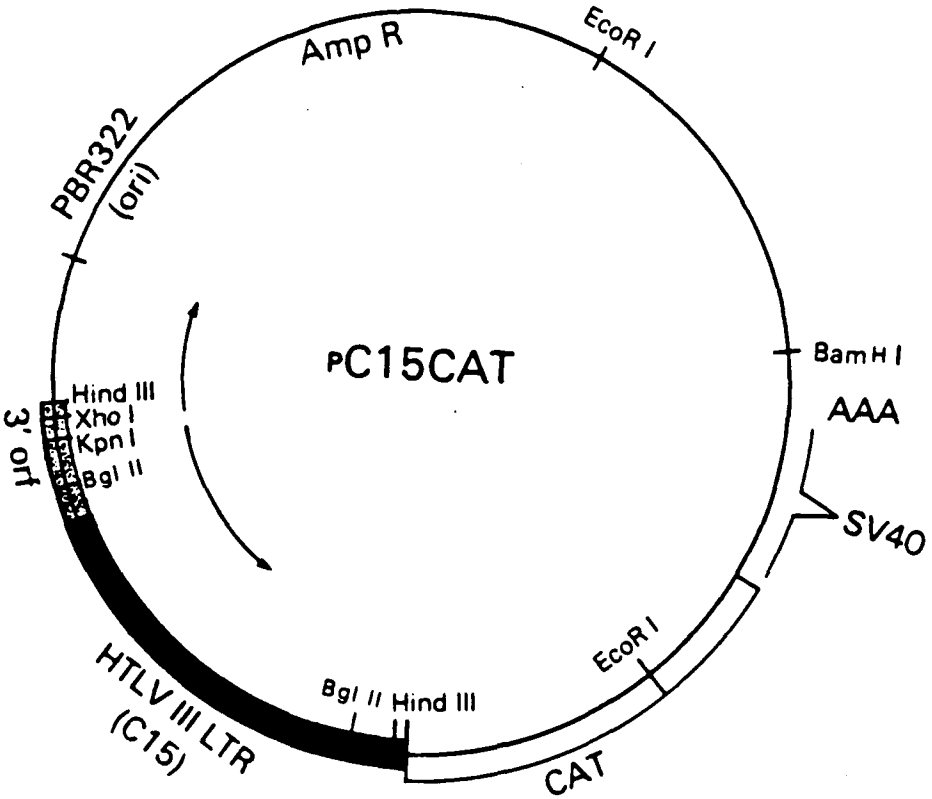
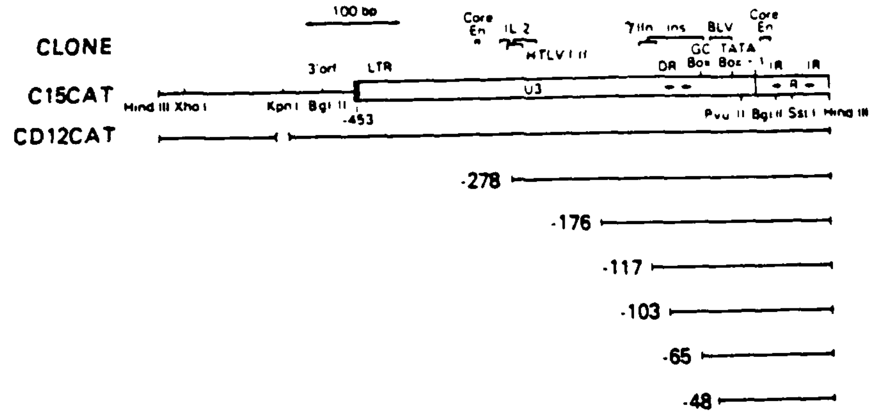


Fig. 4 Plasmid clone C15CAT was constructed by ligation of HindIII linkers to the blunt ended PstI insert of C15, digestion with HindIII and ligation of the resulting fragment into the HindIII site of pSVOCAT. Deletion mutants in LTR were made by Bal31 exonuclease digestion from the KpnI site in the direction of arrows.



CLONE	H9 (RSV = 61 0)*	H9/HTLVIII (RSV = 0 5 ± 0 1)
C15CAT	0 3*	5 1 ± 0 5
CD12CAT	2.0*	39 4 ± 5 9
-278	2.6 ± 0 3	58 0 ± 9 8
-176	2.1 ± 0 3	60 5 ± 12 3
-117	1 1 ± 0.1	55 9 ± 3 2
-103	0.3 ± 0 1	76 2**
-65	0 0	1 6 ± 0 5
-48	0 0	0 0

```
3' orf--->                            Xho I
CCAGCAGCAG ATGGGGTGGG AGCAGTATCT CGACACCTAG AAAAACATGG AGCAATCACA  8505

AGTAGCAACA CAGCAGCTAC CAATGCCGCT TGTGCTTGGC TAGAAGCACA AGAGGAGGAG  8565
                                 Kpn I
GAGGTGGGTT TTCCAGTCAC ACCTCAGGTA CCTTTAAGCC CAATGACTTA CAAGGCAGCT  8625
  Bgl II                 Polypurine Tract  |(-453)U3--->
GTAGATCTTA GCCACTTTTT AAAAGAAAAG GGGGGACTGG AAGGGCTAAT TCACTCCCAA  8685

CGAAGACAAG ATATCCTTGA TCTGTGGATC TACCACACAC AAGGCTACTT CCCTGATTGG  8745
                                                   Core En
CAGAACTACA CACCAGGGCC AGGAGTCAGA TATCCACTGA CCTTTGGATG GTGCTACAAG  8805
      IL-2                HTLV I,II |-278  IL-2
CTAGTACCAG TTGAGCCAGA TAAGGTAGAA GAGGCCAACA AAGGAGAGAA CACCAGCTTG  8865

TTACACCCTG TGAGCCTGCA TGGAATGGAT GACCCGGAGA GAGAAGTGTT AGAGTGGAGG  8925
                |-176
TTTGACAGCC GCCTAGCATT TCATCACATG GCCCGAGAGC TGCATCCGGA GTACTTCAAG  8985
<--3' orf          INS
 -IF  ***       |-117          |-103 Direct Repeats
AACTGCTGAC ATCGAGCTTG CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT  9045
                                              TATA
  |-65                |-48                    Box    Pvu II
GGCCTGGGCG GGACTGGGGA GTGGCGAGCC CTCAGATGCT GCATATAAGC AGCTGCTTTT  9105
                     Core En
   <---U3 |(+1)R--->  IR        Bgl II            Sst I IR
TGCCTGTACT GGGTCTCTCT GGTTAGACCA GATCTGAGCC TGGGAGCTCT CTGGCTAACT  9165
                         Poly A Sig.
                             Hind III            | poly A
AAGGAACCCA CTGCTTAAGC CTCAATAAAG CTTgccttga gtgcttca               9213
```

Fig. 5 Features of clone C15 and resulting 5' deletion clones containing HTLV-III-LTR sequences. CD12 contained a small deletion around the KpnI site by restriction enzyme analysis while retaining the surrounding XhoI and BglII sites. Each clone is numbered corresponding to the distance in nucleotides from the transcriptional start site at +1. Regions of homology to other genes and biologically relevant features are diagrammed as indicated above the map. CAT assay results of lysates prepared from transfected H9 and H9/III cells are shown on the right of each clone. Values are percent conversion of ^{14}C chloramphenical to acetylated metabolites.

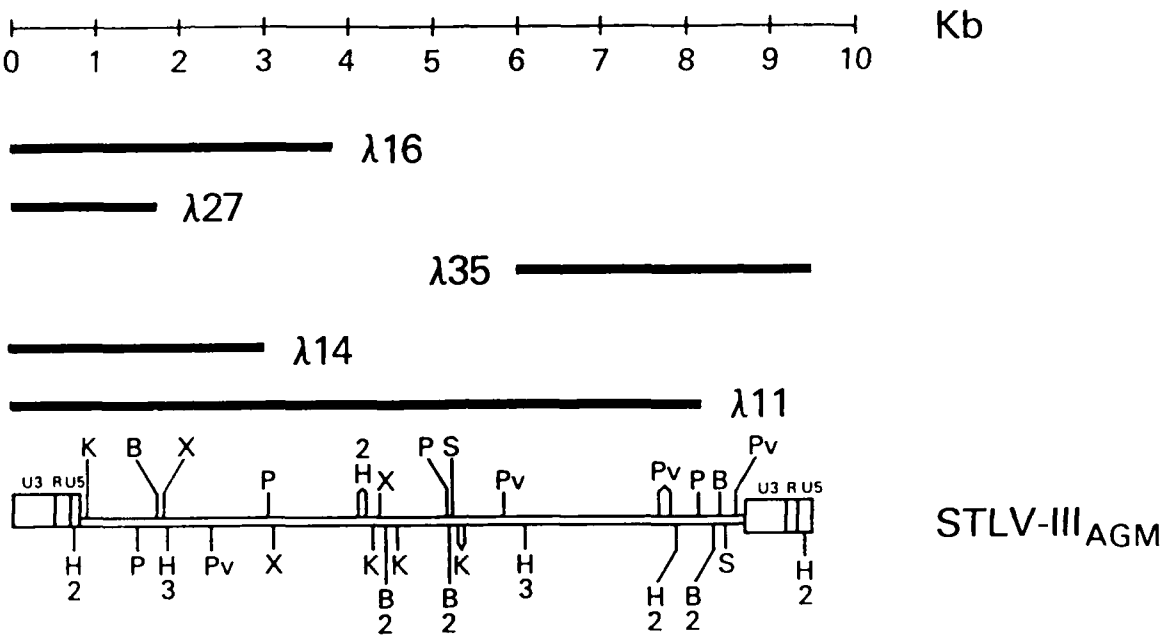
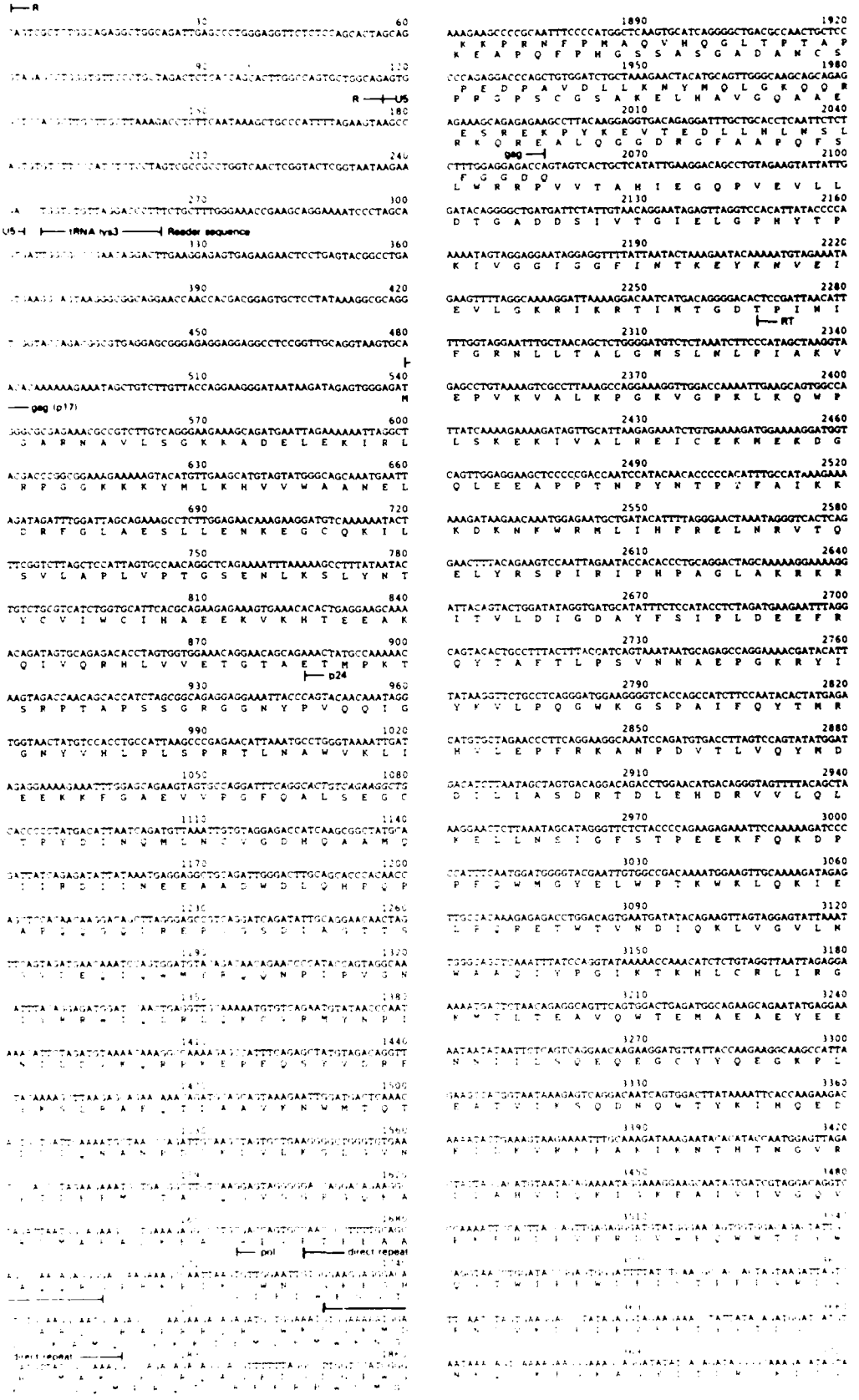
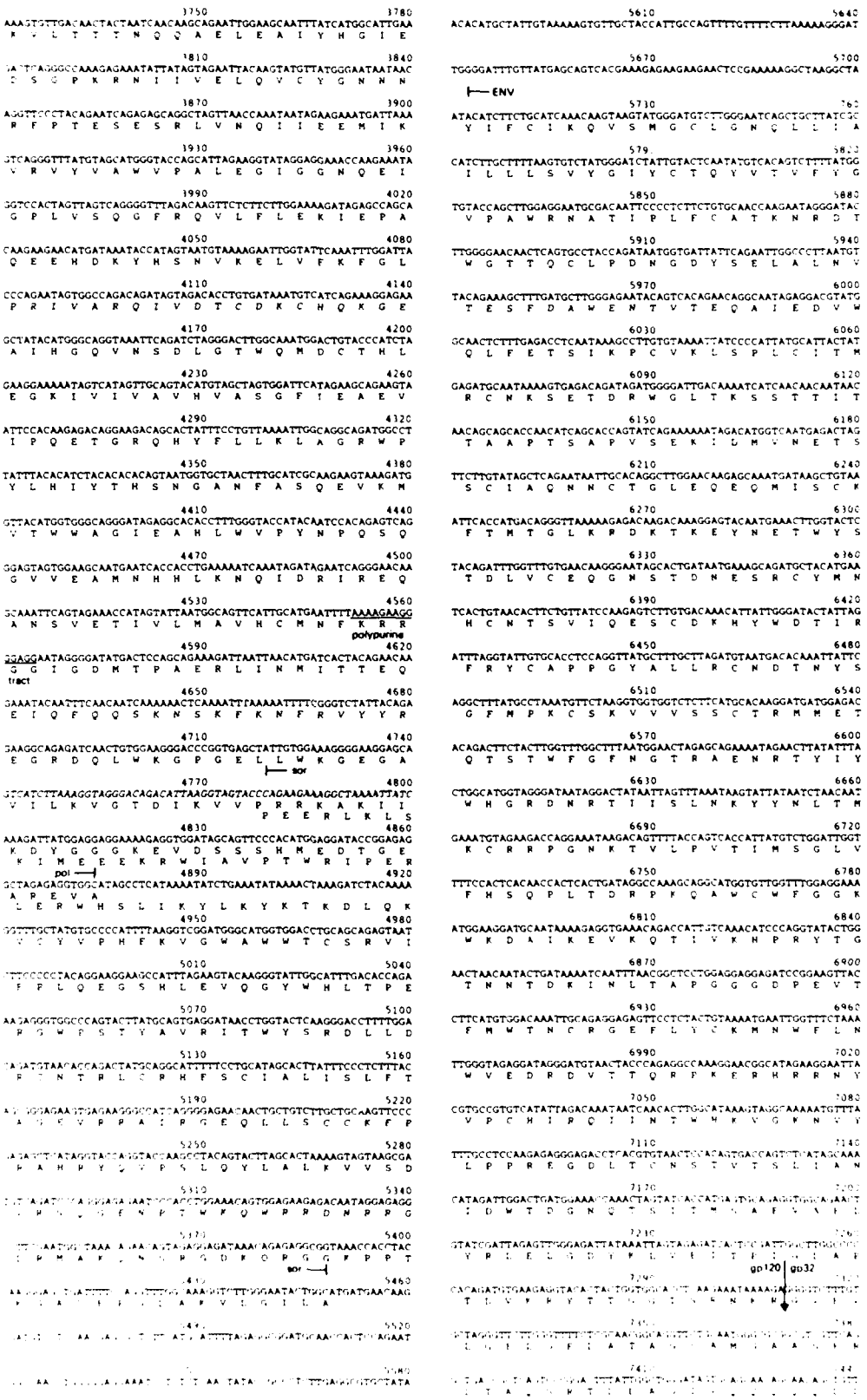


Fig. 6 Restriction enzyme map of the STLV-III$_{AGM}$ genome.

The solid bars in the figure represent the viral content of the lambda clones isolated from the genomic library obtained from the STLV-III$_{AGM}$ infected cell line K6W. The endonuclease map depicted in the lower part of the figure was derived by computer analysis of the DNA sequence obtained from four phage clones 16, 27, 35, and 11 (See Figure 7).


R
U5
tRNA lys3
Reader sequence
gag (p17)
gag
pol
RT
p24
direct repeat
direct repeat


ENV
polypurine
tract
pol
sor
sor
gp120
gp32

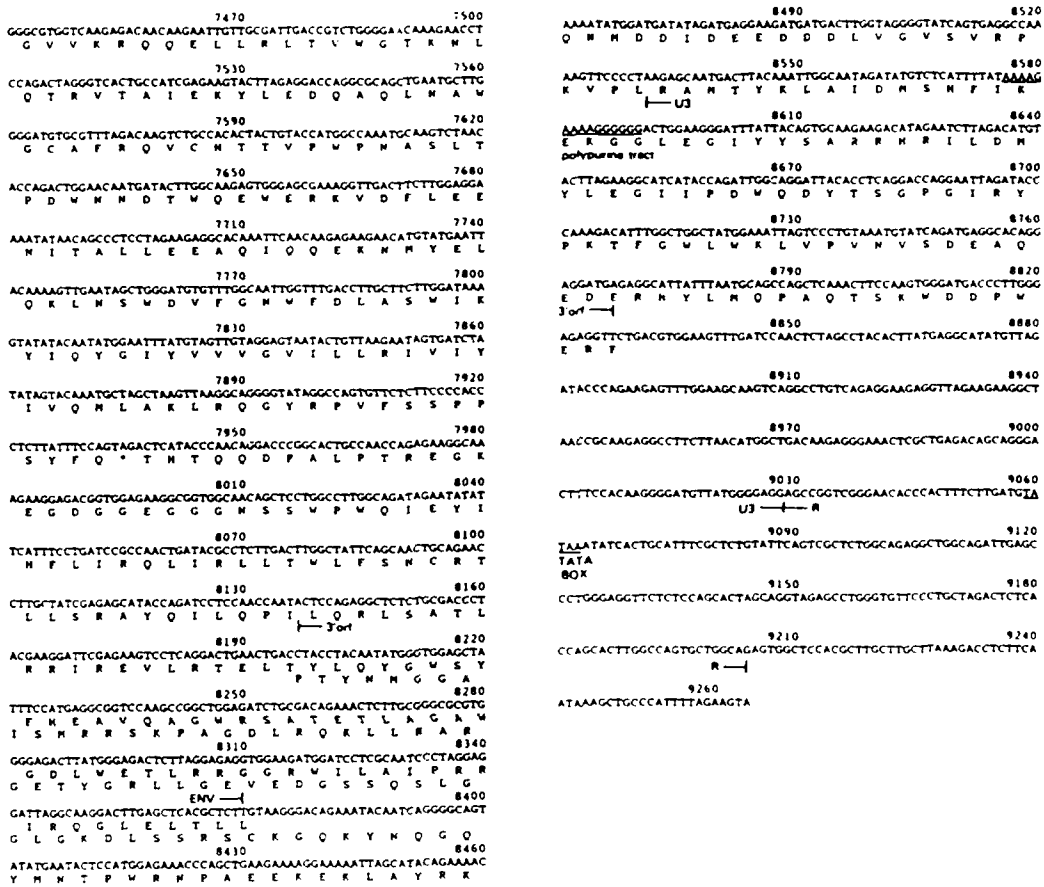


Fig. 7 The complete nucleotide sequence of STLV-III$_{AGM}$
The DNA sequence of the complete provirus was determined with the dideoxy chain termination method of Sanger on single and double stranded DNA. The sequence has been numbered from the beginning of the R region at the 5' end of the provirus to the end of the R region at the 3' end of the proviral DNA. The major gene regions are indicated throughout the length of the DNA sequence.

COMPARISON OF THE GENOMIC ORGANIZATION OF THE STLV-III AND HIV-2_{ROD} PROVIRUSES

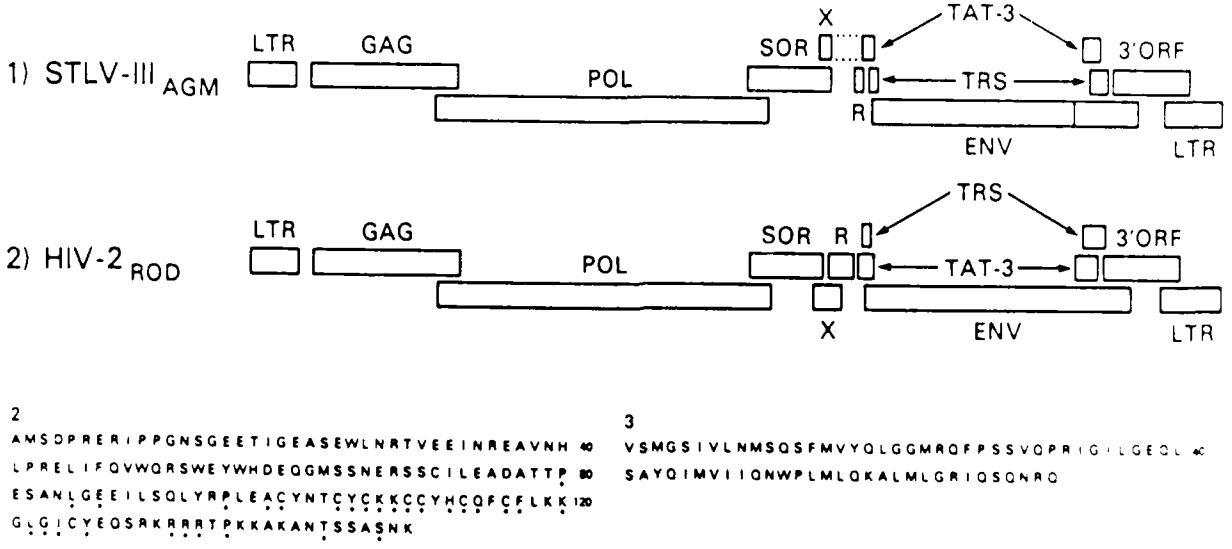


Fig. 8 Genomic Structure of STLV-III_{AGM} and HIV-2_{rod}

A diagrammatic representation of the genomic structure of the STLV-III_{AGM} and HIV-2_{rod} proviruses is shown. In the lower panel, the open reading frames identified in the middle of the genome are given in single-letter code. The numbers 2 and 3 represent the second and third open frames, respectively. The deletion in STLV-III_{AGM} which was found by comparison of the DNA sequences with HTLV-IV and HIV-2_{rod} is shown.

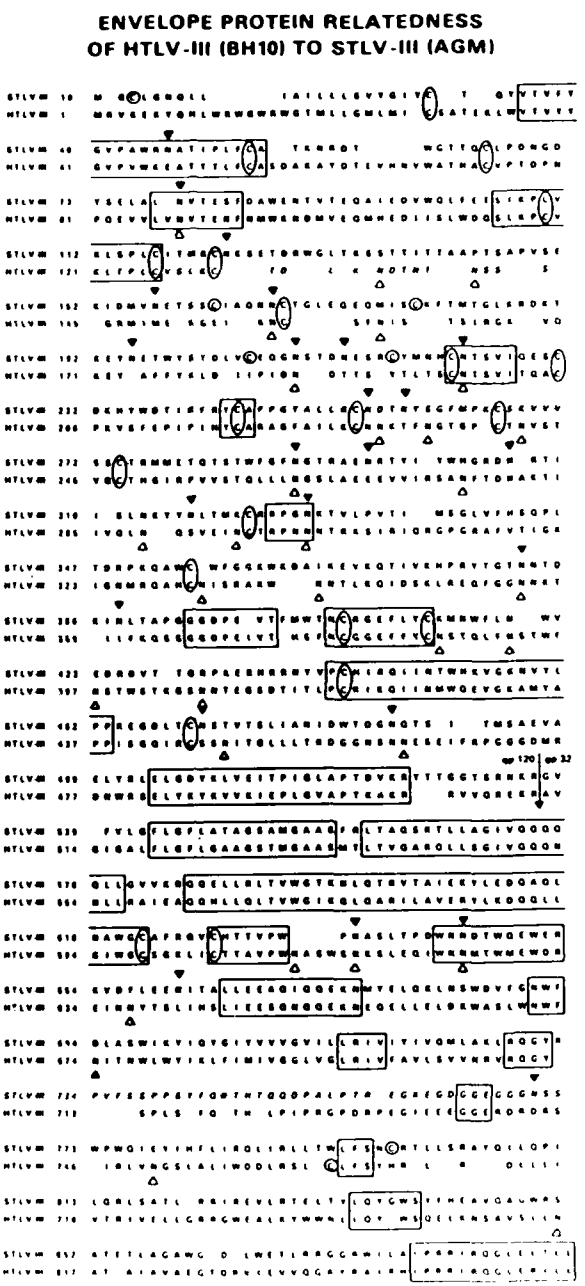


Fig. 9 Similarities between envelope proteins of HTLV-IIIB and STLV-III$_{AGM}$.

The alignment of the derived envelope proteins of STLV-III$_{AGM}$ and HTLV-III$_{B}$ was derived using the protein alignment program of microgene. The dividing arrow indicates the putative cleavage site between the extracellular and transmembrane portions of the precursor envelope protein. The boxes include the most conserved region between STLV-III$_{AGM}$ and HTLV-IIIB envelopes. Circles indicate the location of cysteines in the envelope proteins.

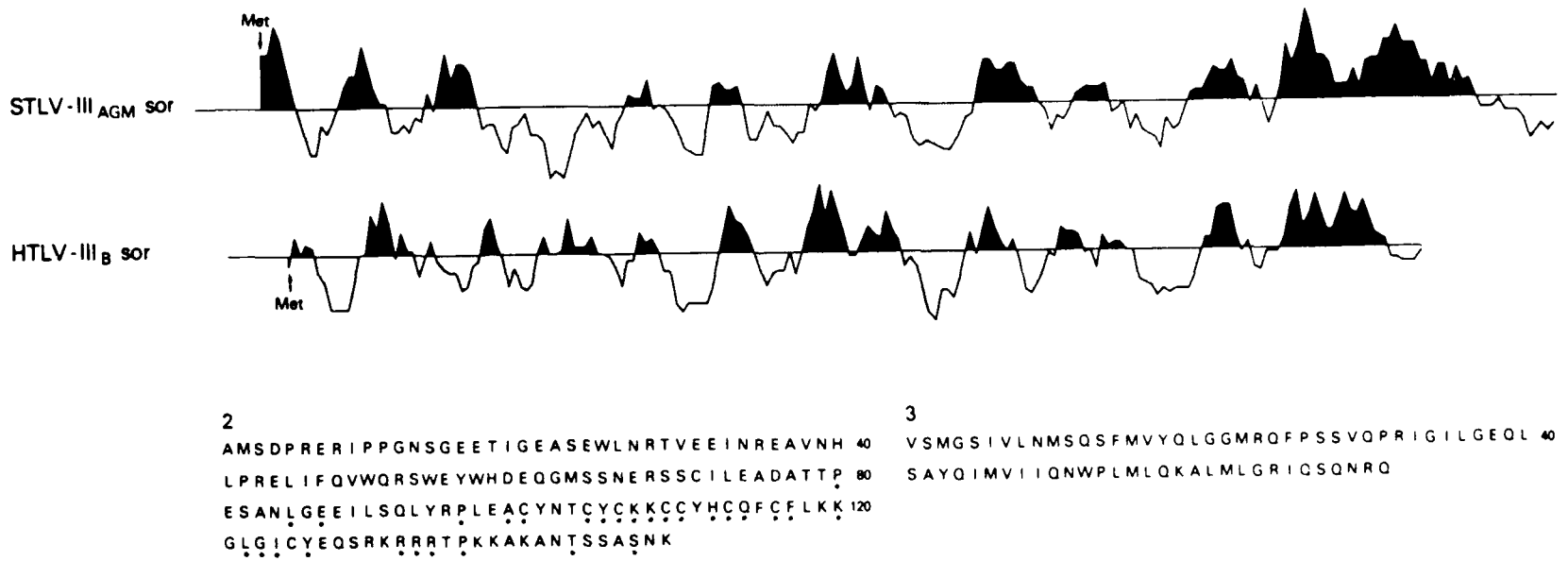


Fig. 10 Hydropathy profiles of 3'ord regions of STLV-III and HIV isolates.

A region of highly conserved amino acid sequence within the 3' open reading frame of STLV-III$_{AGM}$, HIV isolates and HIV-2$_{rod}$ is shown in the above figure. The upper part shows the similar hydropathy profile of the STLV-III$_{AGM}$ and HTLV-IIIB and HIV-2$_{rod}$ proteins. The lower portion of this figure shows the amino acid sequence of STLV-III$_{AGM}$ in the same region which is compared to the other HIV-1 and HIV-2$_{rod}$ isolates.

→R 30 60
AGT[illegible]A[illegible]GGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGG

90 120
[illegible]TGCTA[illegible]CCA[illegible]TGTTGG[illegible]GGCGCTGGGCAGACGG

R ◀|▶ U5
150 180
[illegible]AATAAA[illegible]TGC[illegible]AATTAGAAG[illegible]AA[illegible]TTAAGT

210 240
[illegible]TAGTCGCCGCCTGGTCATCGGTGTTCACCTGTATAACAAGA

270 300
[illegible]TTTGGGAAACCGAGGCAGGCGGAAAATCCCTAGC

U5 ◀| 330 360
AG[illegible]GAACAGGA[illegible]TTGAAGAAGACTGAGAAGCCTTGGAACAAGGCTGGGTG
PBS

390 420
AAGGCAGTAAGGGCGGCAGGAACAAACCAGCACGGAGTGCTCCTAGAAAGGCGCGGGCCG

450 480
AGGTACCAAAG[illegible]GGCGTGTGGACGGGAGTCGAGAGGCCTCCGGGTGAAGGGTAAGTACCT

510 540
ACACCAAAAACTGTAGCCAAAGAGGGCTTGTTATCCTACCTTTAGACAGGTAGAAGATTG

|▶GAG
570 600
TGGGAGATGGGCGCGAGAAACTCCGTCTTGAGAGGGAAAAAAGCAGATGAATTAGAAAAA
M G A R N S V L R G K K A D E L E K

630 660
ATTAGGTTACGGCCCGGCGGAAAGAAAAAGTACAAGTTAAAACATATTGTGTGGGCAGCG
I R L R P G G K K K Y K L K H I V W A A

690 720
AATGAATTGGACAGATTCGGATTAGCAGAGAGCCTGTTGGAGTCAAAAGAAGGTTGCCAA
N E L D R F G L A E S L L E S K E G C Q

750 780
AAAATTCTTACAGTTTTAGATCCATTAGTACCGACAGGTTCAGAAAATTTAAAAAGCCTT
K I L T V L D P L V P T G S E N L K S L

810 840
TTTAATACTGTCTGCGTCATCTGGTGCATACACGCAGAAGAGAAAGTGAAAGATACTGAA
F N T V C V I W C I H A E E K V K D T E

870 900
GGAGCAAAACAAATAGTACAGAGACATCTAGTGGCAGAAACAGGAACTGCAGAGAAAATG
G A K Q I V Q R H L V A E T G T A E K M

930 960
CCAAATACAAGTAGACCAACAGCACCACCTAGCGGGAAAAACTTCCCCGTGCAGCAAGTA
P N T S R P T A P P S G K N F P V Q Q V

990 1020
GCTGGCAACTATACCCATATACCACTGAGTCCCGGAACTCTAAATGCTTGGGTAAAATTA
A G N Y T H I P L S P G T L N A W V K L

1050 1080
GTGGAGGAAAAGAAATTCGGGGCAGAAGTAGTGCCAGGATTTCAGGCACTCTCAGAAGGC
V E E K K F G A E V V P G F Q A L S E G

1110 1140
TGCACGCCCTATGATATTAATCAGATGCTTAATTGTGTGGGCGACCATCAAGCAGCTATG
C T P Y D I N Q M L N C V G D H Q A A M

1170 1200
CAAATAATCAGGGAAATTATTAATGAGGAAGCAGCAGACTGGGATGTTGCACATCCCATA
Q I I R E I I N E E A A D W D V A H P I

1230 1260
CCAGG[illegible]TTA[illegible]CAGGGGGCAGCTTAGAGAGCCAAGAGGGTCTGACATAGCAGGGACA
P G P L P A G Q L R E P R G S D I A G T

1290 1320
[illegible]CAGTGGATGTTTAGACCACAAAATCCTGTACCAGTA
[illegible] Q W M F R P Q N P V P V

1350 1380
[illegible]CAGATAGGA[illegible]TGGCGTCAGGATGTACAAT
[illegible] R W I Q I G L Q K C V R M Y N

1410 1440
[illegible]
[illegible]

1470 1500
[illegible]
[illegible]

1530 1560
[illegible]
[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1830 1860
GTCATGGCAAATTGCCCAGATAGACAGGCTGGTTTTTTAGGGATTGGCCACTGGGGAAAG
V M A N C P D R Q A G F L G I G H W G K
pol ▶ T G W F F R D W P L G K

1890 1920
AAGGCCCGCAACTTCCCCGCGGCCCAAGTCCCGCAGGGGCTAACACCAACAGCACCCCCA
K A R N F P A A Q V P Q G L T P T A P P
E G P Q L P R G P S P A G A N T N S T P

1950 1980
TTGGATCCAGCAGTGGACCTACTGGAGAAATATATGCAGCAAGGAAAAAGGCAAAGGGAG
L D P A V D L L E K Y M Q Q G K R Q R E
I G S S S G P T G E I Y A A R K K A K G

2010 2040
CAGAGAGAGAGACCGTACAAGGAAGTGACAGAGGACTTACTGCGTTTCGAGCAGGCAGAG
Q R E R P Y K E V T E D L L R F E Q A E
A E R E T V Q G S D R G L T A F R A G R

2070 2100
ACACCATGCAGGGAGACGACAGAGGACTTGCTGCACCTCAATTCTCTCTTTGGAAAAGAC
T P C R E T T E D L L H L N S L F G K D
D T M Q G D D R G L A A P Q F S L W K R

2130 2160
CAGTAGTCACAGCACACATTGAGGGTCAGCCAGTAGAAGTTTTGTTAGACACAAGGGCTA
P V V T A H I E G Q P V E V L L D T R A

2190 2220
ACGACTCAATAGTAGCAGGAATAGAGTTAGGGAGCAATTATAGTCCAAAAATAGTAGGAG
N D S I V A G I E L G S N Y S P K I V G

2250 2280
GAATAGGGGGATTCATAAATACCAAGGAATATAAAAATGTAGAAATAGAAGTCCTAGGTA
G I G G F I N T K E Y K N V E I E V L G

2310 2340
AAAGGGTAAAAGCCACCATAATGACAGGTGATACCCCGATCAACATTTTTGGCAGAAATG
K R V K A T I M T G D T P I N I F G R N

2370 2400
TTCTGACAGCTTTAGGCATGTCATTAAACCTGCCAGTTGCCAAGATAGAACCAATAAAA
V L T A L G M S L N L P V A K I E P I K

2430 2460
TAATGCTAAAGCCAGGGAAAGATGGACCAAGACTAAAACAATGGCCTTTAACAAAAGAAA
I M L K P G K D G P R L K Q W P L T K E

2490 2520
AAATAGAAGCACTAAAAGAAATCTGTGAAAAAATGGAAAAAGAAGGCCAGCTAGAAGAAG
K I E A L K E I C E K M E K E G Q L E E

2550 2580
CACCTCCAACTAATCCTTATAATACCCCCACATTTGCAATCAAGAAAAAGGACAAAAACA
A P P T N P Y N T P T F A I K K K D K N

2610 2640
AATGGAGAATGCTAATAGACTTTAGGGAGCTAAACAAGGTAACTCAGGATTTCACAGAAA
K W R M L I D F R E L N K V T Q D F T E

2670 2700
TTCAGTTAGGAATTCCACACCCAGCAGGATTGGCCAAAAAGAGGAGAATTACTGTACTAG
I Q L G I P H P A G L A K K R R I T V L

2730 2760
ATGTAGGGGATGCTTACTTTTCCATACCACTACATGAGGACTTTAGACAGTATACTGCAT
D V G D A Y F S I P L H E D F R Q Y T A

2790 2820
TCACTCTACCATCAGTAAACAATGCAGAACCAGGAAAAAGATATATATACAAAGTCCTAC
F T L P S V N N A E P G K R Y I Y K V L

2850 2880
CACAGGGGTGGAAGGGGTCACCAGCAATTTTTCAATACACAATGAGGCAGATCTTAGAAC
P Q G W K G S P A I F Q Y T M R Q I L E

2910 2940
CATTCAGAAAAGCAAACGAGGATGTCATTATCATTCAGTACATGGATGATATCTTAATAG
P F R K A N E D V I I I Q Y M D D I L I

2970 3000
CTAGCGACAGGACAGACTTAGAACATGACAAAGTGGTCCTGCAGCTAAAAGAACTTCTAA
A S D R T D L E H D K V V L Q L K E L L

3030 3060
ATGGACTAGGATTTTCCACCCCAGATGAGAAATTCCAGAAAGACCCTCCATATCGCTGGA
N G L G F S T P D E K F Q K D P P Y R W

3090 3120
TGGGCTATGAATTATGGCCAACTAAATGGAAGTTGCAAAAAATACAGCTGCCCCAAAAAG
M G Y E L W P T K W K L Q K I Q L P Q K

3150 3180
[illegible]
E V W T V N D I Q K L V G V L N W A A Q

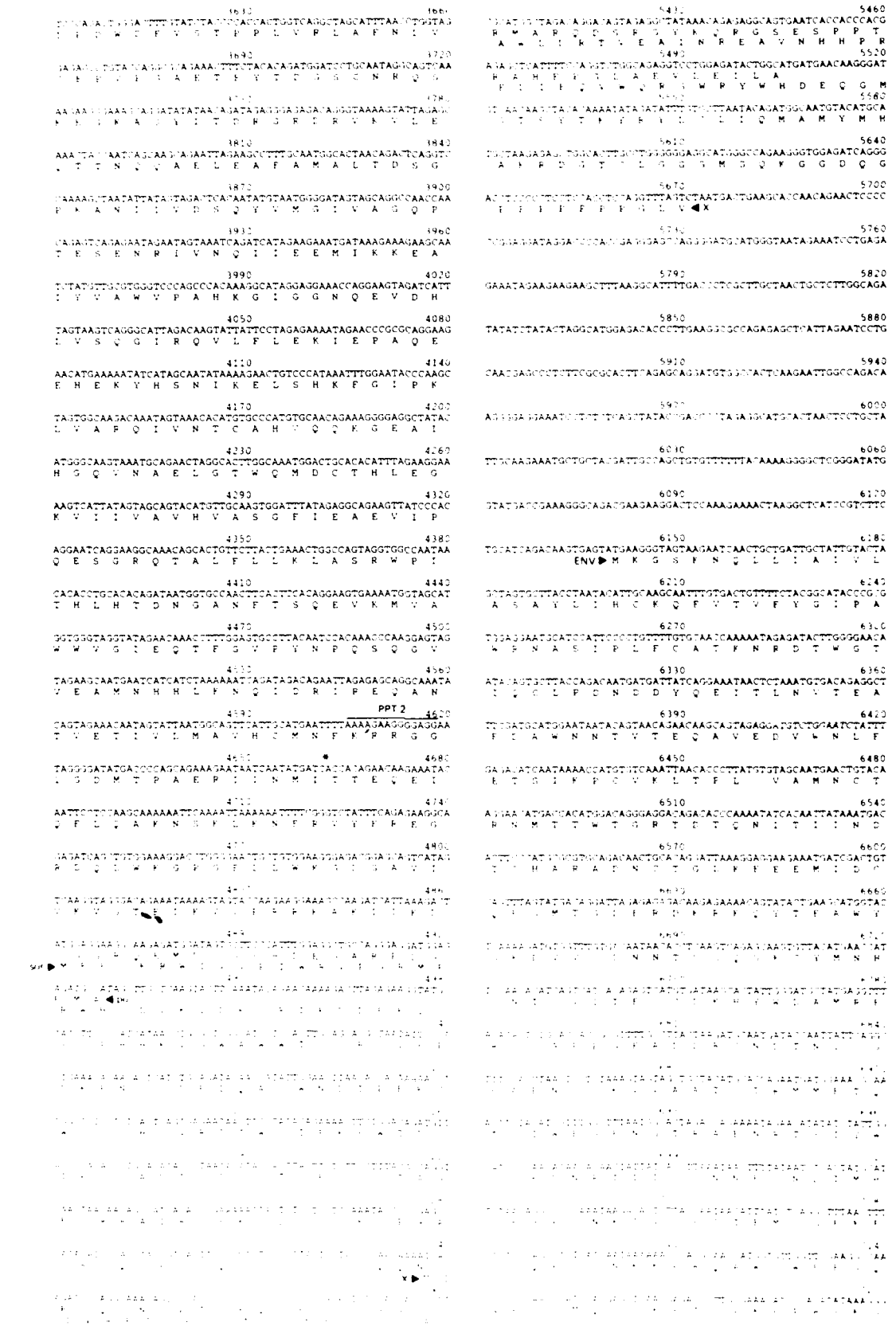
3210 3240
[illegible]
I Y P G I K T K H L C R L I R G K M T L

[illegible]

[illegible]

[illegible]

[illegible]



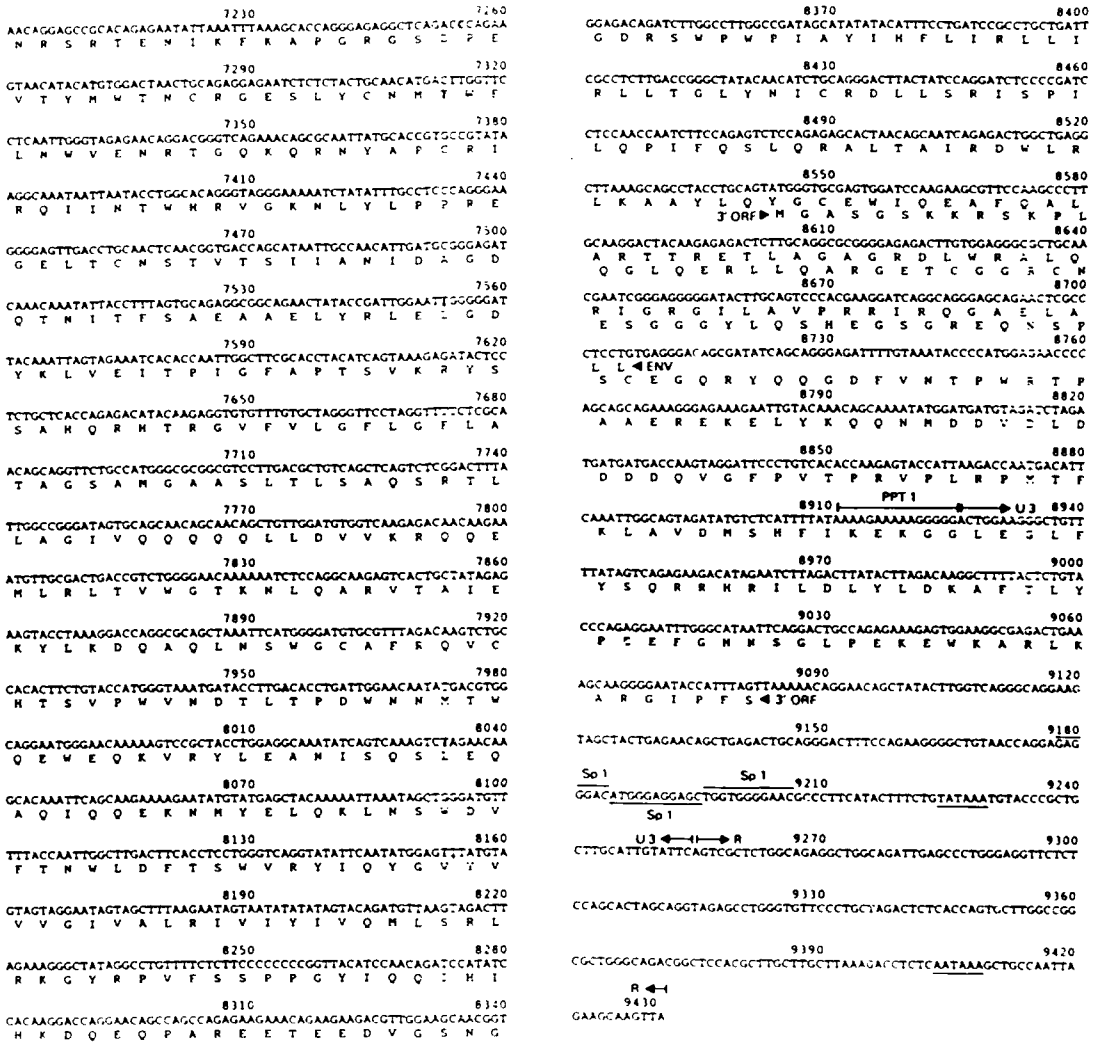


Fig. 11 The complete nucleotide sequence of the HIV-2_{NIH-Z} virus.

The DNA sequence has been organized and presented as the U5-R-gag-pol-middle region-env-3'orf-U3-R. The translation in amino acids of all major open reading frames is presented underneath the nucleotide sequence.

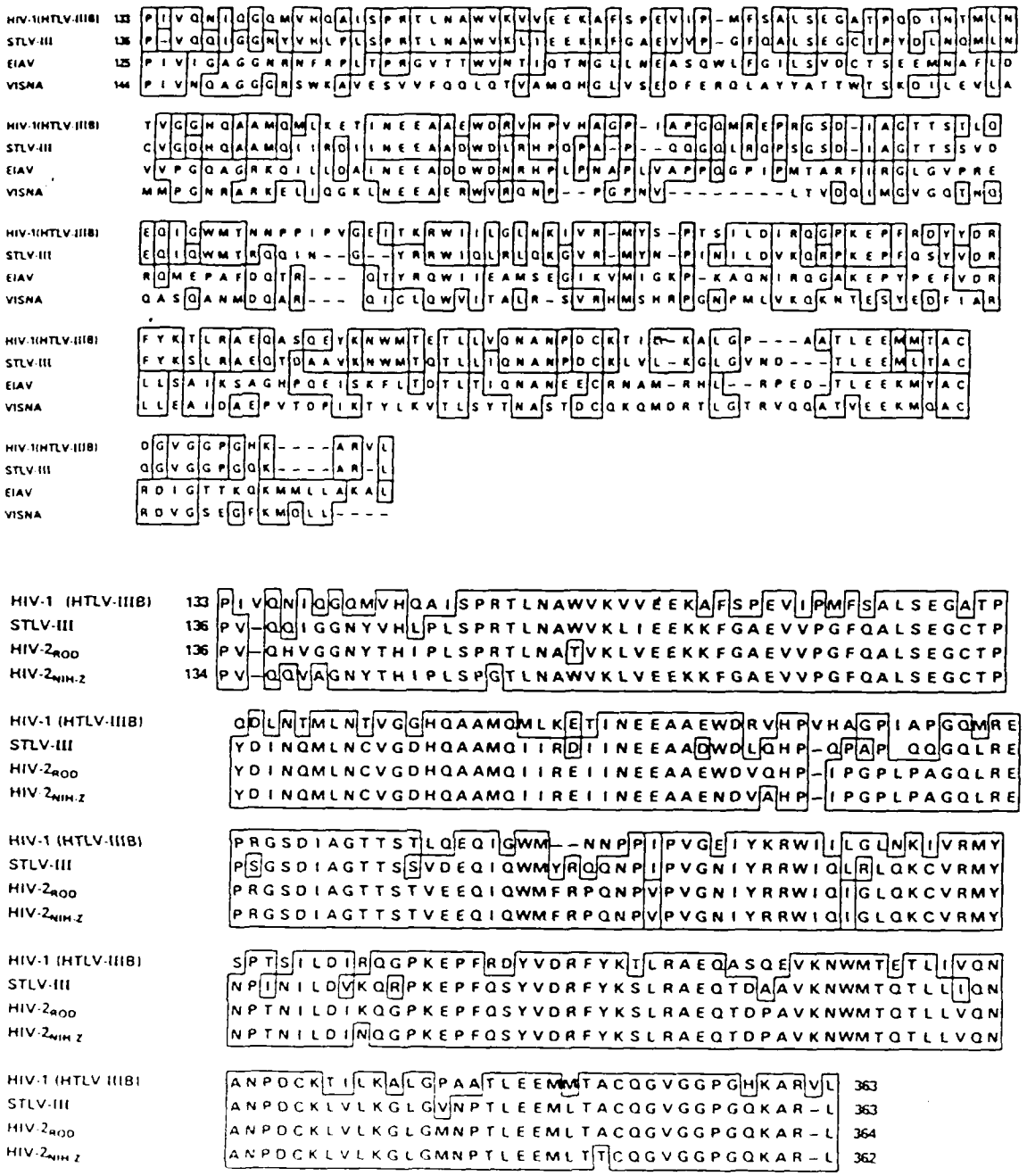


Fig. 12 Comparison of the major core proteins of HIV-1, HIV-2, STLV-III$_{AGM}$, EIAV and VISNA

The upper portion of this figure represents the amino acid alignment in the major core proteins (p24-26) of the HTLV-IIIB strain of HIV-1, STLV-III$_{AGM}$, EIAV and VISNA. The boxed regions indicate amino acids that are identical among all four viral gag proteins. The lower portion of this figure represents a similar analysis of the major core proteins of STLV-III, both HIV-2 isolates and HIV-1 (HTLV-IIIB).

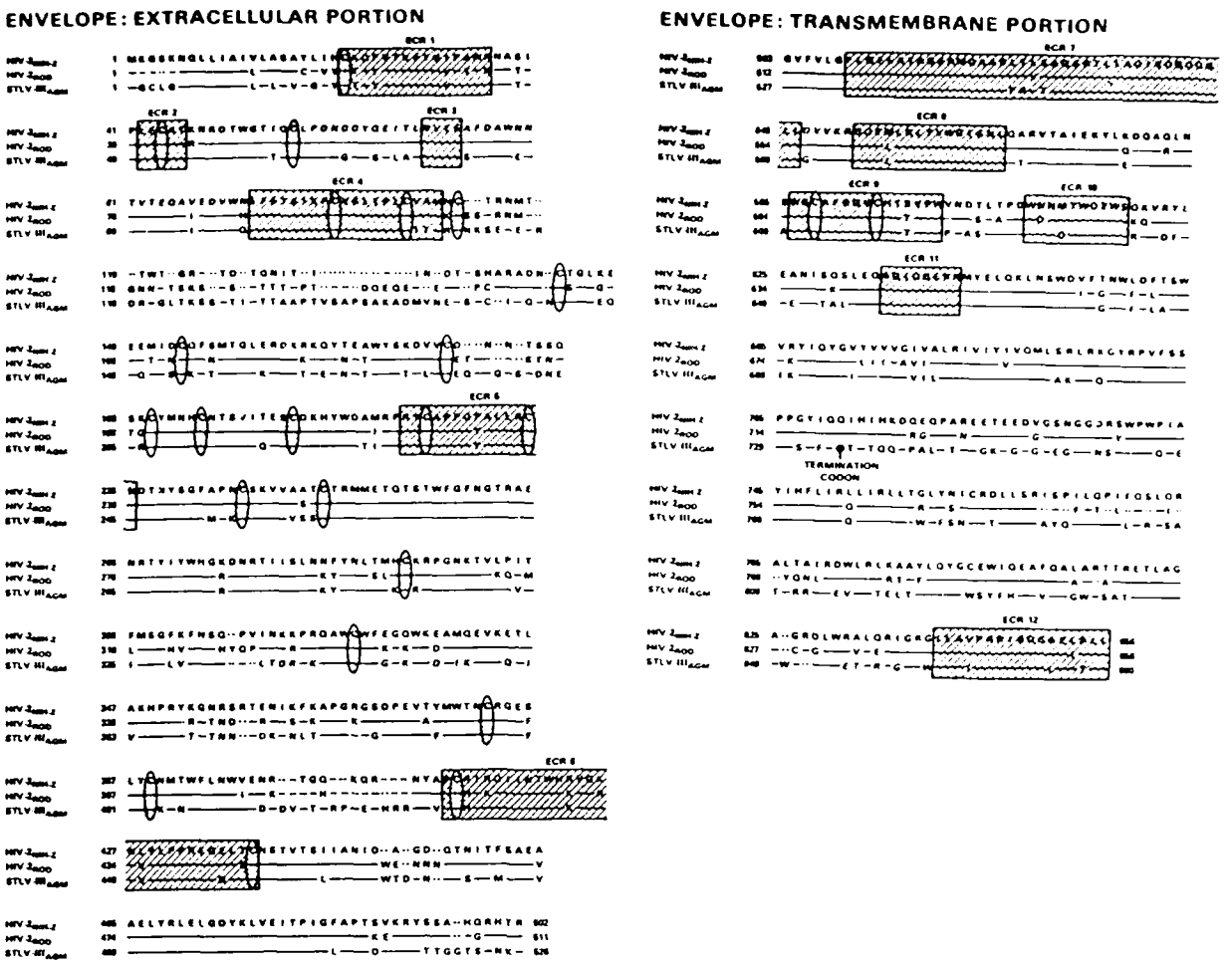


Fig. 13 Identification of variable and conserved regions in the envelope proteins of HIV-1 and HIV-2.

This figure represents a comparative analysis of the envelope proteins of HIV-2_{NIH-Z}, HIV-2_{rod} and STLV-III_{AGM} by amino acid alignment. The left portion of this figure represents the extracellular envelope proteins while the right portion contains the transmembrane portion of the envelope proteins. The amino acid sequence of the HIV-2_{NIH-Z} envelope gene is reported as a line indicating a perfect match in amino acid sequence while the dotted line stands for the lack of the amino acid. The boxes include regions (ECR: envelope conserved regions) which match well and exhibit only conservative amino acid changes with respect to the HTLV-IIIB strain of HIV-1. The dotted ovals indicate highly conserved cysteine residues in both HIV-2 isolates, STLV-III_{AGM} and HIV-1 (HTLV-IIIB). The empty ovals indicate cysteine residues that are conserved only in the West African viral isolates and STLV-III_{AGM}. The location of the termination codon in the STLV-III_{AGM} transmembrane envelope protein is indicated.

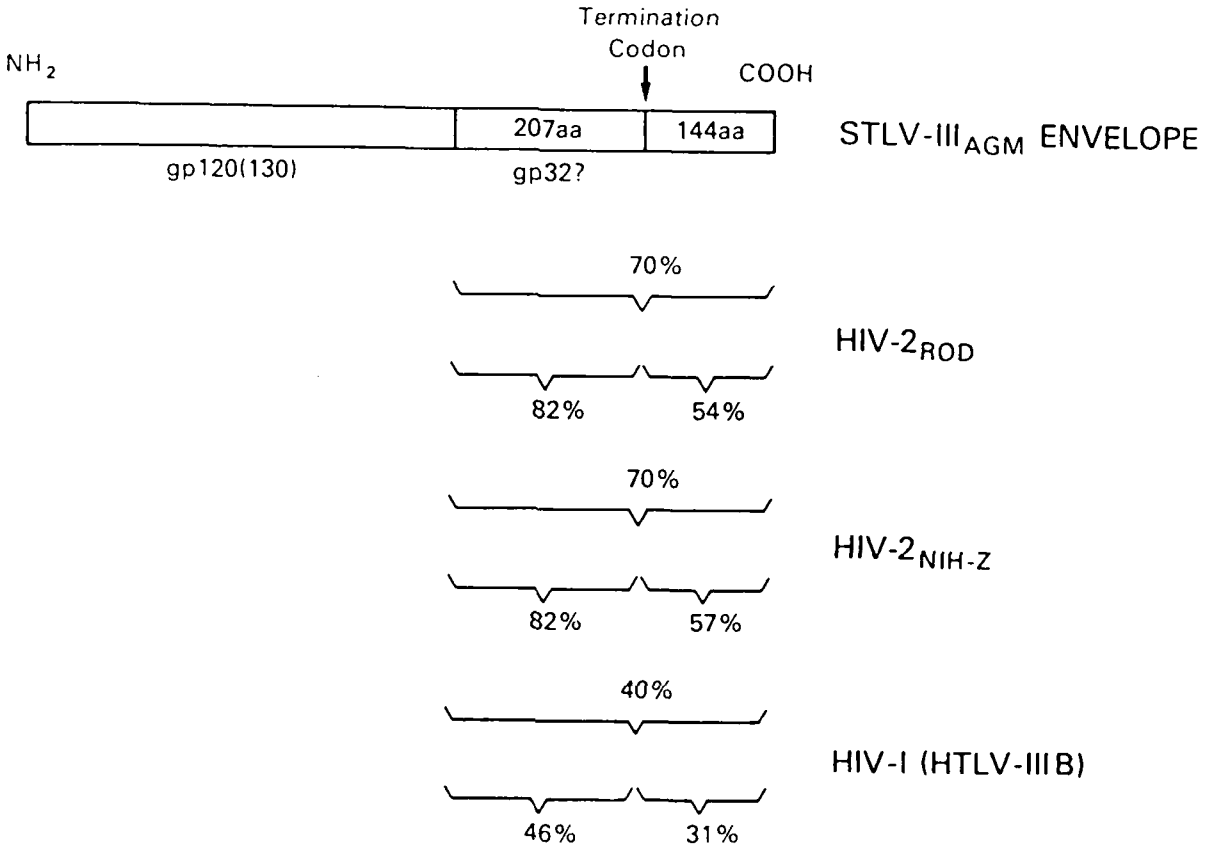


Fig. 14 Schematic representation of the envelope proteins of STLV-III_{AGM} and percentage of amino acid conservation in the other viral isolates before and after the stop codon.

A schematic representation of the transmembrane region of the envelope protein is presented and the overall percentage of homology among STLV-III_{AGM} and both HIV-2 and HIV-1 is provided. Underneath the line, the homology between each viral transmembrane protein and the STLV-III_{AGM} transmembrane protein is represented before and after the stop codon.

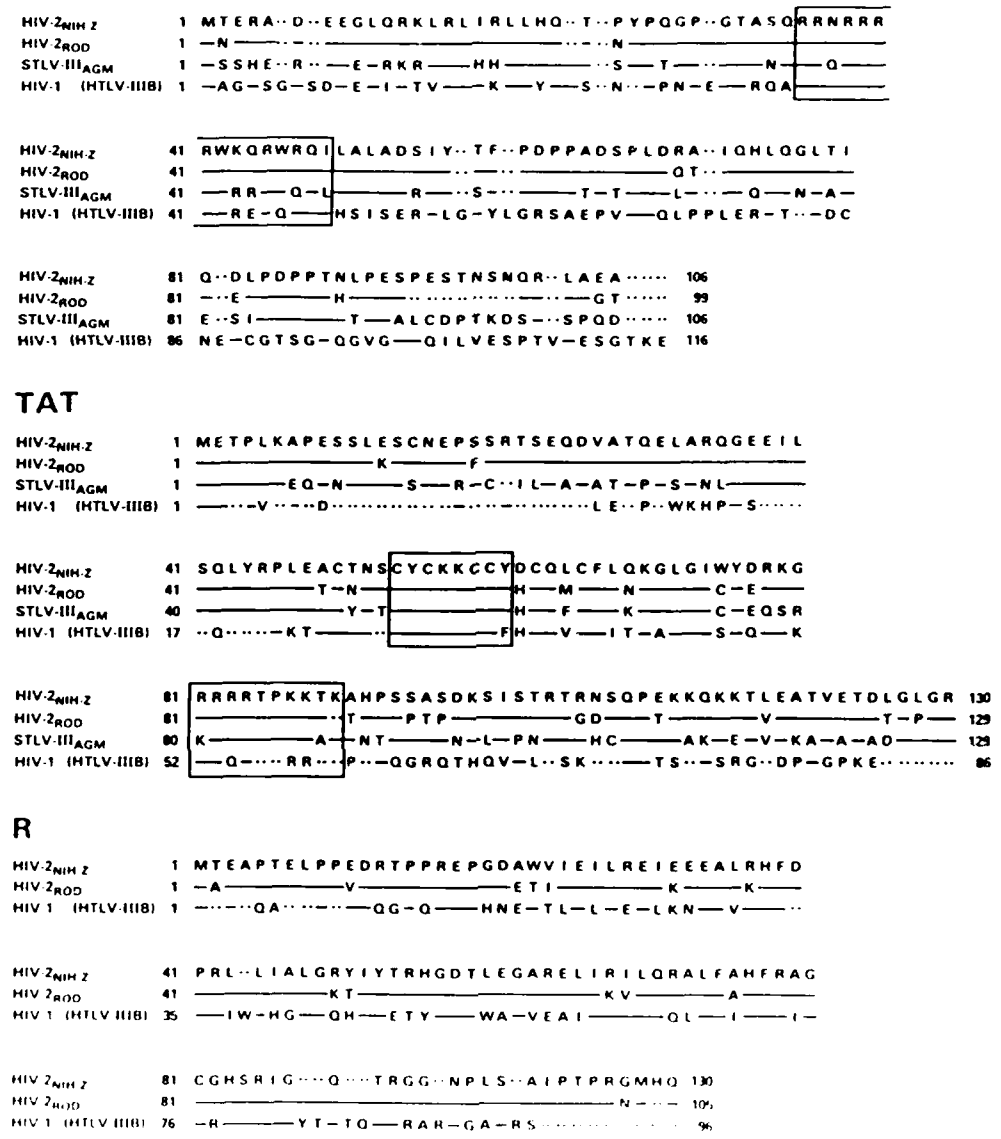


Fig. 15 Amino acid alignment of other viral proteins.

The amino acid sequence of the HIV-2_{NIH-Z} trs, tat and R presumed proteins is represented in the top line of each panel. The continuous line represents matching amino acids among the viral isolates. The boxes include regions which are conserved in HIV 1 (HTLV-IIIB).

Sample	Supernatants removed from Molt 3 cultures and concentrated x 100		Infectivity	Trans-activation in H9 cells	Trans-activation in Cos-1 cells
	RT Activity dpm	Virus Concentration (particles per milliliter)		Mean Conversion (%)	Mean Conversion (%)
X	138,857	3.7×10^{10}	+	40.5 ± 3.9	56.1 ± 19.9
S	137,743	2.6×10^{10}	-	32.0 ± 6.9	67.5 ± 28.9
3.3	120,262	2.9×10^{10}	-	43.0 ± 8.9	62.0 ± 17.5
6.9	74,794	2.4×10^{10}	-	54.3 ± 23.9	37.5 ± 15.5
153	61,383	3.3×10^{10}	-	29.8 ± 2.5	47.0 ± 25.7

TABLE 1: Infectivity and properties of sor mutant viruses of HIV 1. M t 3 were infected with HIV 1 by coculturing with Cos-1 cells transfected with X, S, 3.3, 6.9 and 153 virus. The presence of virus was assayed by Reverse Transcriptase activity. Viral particles were counted by electron-microscopy. Transactivation activity was measured by co-transfecting the plasmid along with pC15CAT into H9 or Cos-1 cells. The resulting CAT activity was determined in each case.

Construct	Amino acid changes in env (gp41)	DELETIONS tat/trs	LTR/PPT	3'orf	REPLICATION & TRANSMISSION Replication (RT/Cos-1)	Transmission Molt3	H9
LR 329	-0 +47			+	+++++	+++++	+++
LR 206	-4 +59			+	+++++	(t)	+++
X10-1	-5 +15			+	+++++	+++++	+++++
X9-3	-5 +>70			+	+++++	(t)	(t)
LR 369*	-6 +2			+	not tested	++	+++
LR 295	-14 +2			+	+++++	++	+++++
LR 372*	-15 +4			+	not tested	(t)	+++
LR 429*	-15 +4			+	not tested	(t)	+++
LR 269	-17 +2			+	++++	(t)	not tested
LR 306	-22 +2			+	not tested	-	not tested
LR 319	-30 +			+	not tested	-	not tested
LR 468*	-33 +0			+	+++	(t)	+++
LR 362	-37 +18			+	+++	-	not tested
LR 358	-6 +3		+	+	++	(t)	-
LR 318	-6 +24		+	+	+++++	(t)	not tested
LR 204	-6 +27		+	+	+++++	(t)	not tested
LR 312	-14 +74		+	+	+++++	-	"
LR 192	-17 +11		+	+	+++++	-	"
LR 189	-41 +0		+	+	+++++	-	"
LR 274	-42 +2		+	+	+++++	-	"
LR 194	-76 +9		+	+	+++++	-	"
LR 360	-87 +32	+	+	+	-	-	not tested
LR 327	-117+20	+	+	+	-	-	"
LR 230	none		?	+	+	+++	not tested
LR 330	none			+	+++++	+++++	"

KEY ** Denotes the presence of linkers in the plasmid construct

'Replication' was assessed as the production of reverse transcriptase in supernatants taken from cos-1 cells, 6-7 days after transfection. 'Transmission' was judged by the rise in HIV-1 positive cells among cultures of either H9 or Molt3 cells, after coculture with transfected cos-1 cells. 't' denotes transient HIV-1 expression in cultures where no infected cells could be detected by week 4.

\+ <.01% expression by week 4
++ >.01-<1% expression by week 4
+++ 1-10% expression by week 4
++++/++++++ 50-90 % expression by week 4

TABLE 2: Summary of Biological Properties of Deletion Mutants in 3'orf.

CAT ASSAY RESULTS IN T-CELLS (H9) AND HTLV-III INFECTED T-CELLS (H9/III)

Plasmid	Description	20 min. assay H9/III Exp. 2	15 hr. assay H9 Exp. 2	15 hr. assay H9 Exp. 3
SV0	Negative Control	0.0	0.0	0.0
SV2	Positive Control	2.6**	6.7 ± 1.0	4.6 ± 1.5
RSV	"	18.8 ± 4.8	48.4 ± 3.41	55.8 ± 3.1
C15CAT	HTLV III LTR-CAT	20.8**	0.1 ± 0.04	ND
CD12CAT	"	88.8 ± 10.9	0.7 ± 0.05	0.70 ± 0.080
VHHCAT	"	92.7 ± 1.3	ND	2.4 ± 0.34
-65	del. -65 from CAP	32.1**	0.09 ± 0.03	0.0
-48	del. -48 from CAP	0.0	ND	0.0
-65E2	5'-3' -105 to -80	99.1 ± 0.1	ND	2.7 ± 0.53
-65E5	3'-5' -80 to -104	99.2 ± 0.1	ND	0.85 ± 0.31
-48E9	5'-3' -105 to -80	26.7**	ND	0.21 ± 0.061
-48E14	5'-3' -105 to -79	22.4**	ND	0.18 ± 0.006
-48E8	3'-5' -80 to -104	53.8 ± 7.9	ND	0.27 ± 0.093
-117	del. -117 from CAP	97.5 ± 0.8	0.4 ± 0.07	0.66 ± 0.14
-117 Δ BS	del. BglII to SstI	0.13 ± 0.02	0.2 ± 0.09	ND
-117 Δ S	del. 4 bp at SstI	1.2 ± 0.3	16.7**	ND

Table 3. Percent conversion of ^{14}C-chloramphenicol to acetylated metabolites using lysates of cells transiently transfected with the indicated plasmids. Cells used were H9, an uninfected T cell line and H9/III cells, the same cell line productively infected with HTLV III. Transfections were done in triplicate for values showing standard deviations. Values for single and averages for duplicate transfections are indicated by one or two asterisks respectively. ND means not done.

TABLE 4: Per Cent Amino-Acid Homologies Between Virus Isolates

gag		pol		env				
p17	p24	protease	RT	large env	small env	sor(Q)	3'orf(F)	Weighted Average
$STLV\text{-}III_{AGM}$ and $HIV\text{-}2_{ROD}$								
82		76		70		64	56	72
84	88	88	80	71	70			
$STLV\text{-}III_{AGM}$ and HIV-1 ($HTLV\text{-}III_B$)								
51		53		34		24	34	43
$HIV\text{-}2_{ROD}$ and HIV-1 $(HTLV\text{-}III)_B$								
52		55		35		28	20	43

TABLE 5: Amino Acid Homology of Viral Proteins Among Viral Isolates

$HIV\text{-}2_{NIH\text{-}2}$	GAG	POL	ENV			SOR	TAT	TRS/ART	X	R
			TOTAL	ECP	TMP					
$HIV\text{-}2_{ROD}$	92	91	80	78	84	86	85	86	78	69
$STLV\text{-}III_{AGM}$	82	75	70	68	72	64	63	59	-	-
HIV-I (HTLV-IIIB)	52	54	35	32	39	28	29	34	-	37

TABLE 6: Amino Acid Homology of the major gag Proteins (P24/26)

	HIV-2_{NIH-2}	HIV-2_{ROD}	STLV-III_{AGM}	HIV-I (HTLV-IIIB)	EIAV	VISN
HIV-2_{NIH-2}	100	96	88	67	29	24
HIV-2_{ROD}	96	100	88	68	28	25
STLV-III_{AGM}	88	88	100	66	29	26
HIV-I (HTLV-IIIB)	67	68	66	100	29	26
EIAV	29	28	29	28	100	29
VISNA	24	25	26	26	29	100

V. REFERENCES

1. Wong-Staal, F., and Gallo, R.C. (1985). Human T-lymphotropic retroviruses. Nature 317, 395-403.

2. Fisher, A.G., Collalti, E., Ratner, L., Gallo, R.C., Wong-Staal, F. (1985) A molecular clone of HTLV-III with biological activity. Nature 316, 262-265.

3. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. Cell 40, 9-17.

4. Ratner, L., Haseltine, W., Paturea, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Antoni Rafalski, J., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberg, J.A., Papas, T.K., Ghrayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985). Complete nucleotide sequence of the AIDS virus HTLV-III. Nature 313, 277-284.

5. Lee, T.H., Coligan, J.E., Allan, J.S., Mclane, M.F., Groopman, J.E., Essex, M. (1986). A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses. Science 231, 1546-1549.

6. Sodroski, J., Goh, W.C., Rosen, C., Tartar, A., Portelle, D., Burny, A., Haseltine, W., (1986). Replicate and cytopathic potential of HTLV-III/LAV with sor gene deletions. Science 231, 1549-1553.

7. Levi, and Seidman (1979) J. Virol. 29, 328-335.

8. Gallis, et al., (1979) Virology 94, 146-161.

9. Linial, et al., (1978) Cell 15, 1371-1381.

10. Mann, R. and Baltimore D. (1985) Varying the position of a Retrovirus Packaging Sequence results in the encapsidation of both unspliced and spliced RNAs. J. Virology 54, 401-407.

11. Fisher, A.G., Feinberg, M.B., Josephs, S.F., Harper, M.E., Marselle, L.M., Reyes, G., Gonda, M.A., Aldorini, A., Debouk, C., Gallo, R.C. (1986). The trans-activator gene of HTLV-III is essential for virus replication. Nature 320, 367-371.

12. Ivanoff, L.A., Towatari, T., Ray, J., Korornt, B.D., and Petteway, S.R., (1986) Expression and site specific mutagenesis of the poliovirus 3C prolease in Escherichia coli. Proc. Natl. Acad. Sci. USA 83, 5392-5396.

13. Graham, F.L., vander Eb., A.J., Heijneker, H.L. (1984) Nature 251, 687.

14. Curran, J.W., Morgan, W.M., Hardy, A.M., Jaffe, H.W., Darrow, W.W. and Dowdle, W.R. (1985) The epidemiology of AIDS: Current status and future prospects. Science 229, 1352-1357.

15. Salahuddin et al. (1985) Proc. Natl. Acad. Sci. USA 82, 5530.

16. Sodroski, J., Rosen, C., Wong-Staal, F., Salahuddin, S.Z., Popovic, M., Arya, S.K., Gallo, R.C., Haseltine, W.A. (1985) Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science 227, 171-173.

17. Arya, S.K., Guo, C., Josephs, S.F., Wong-Staal, F. (1985) Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). Science 229, 69-73.

18. Rosen, C.A., Sodroski, J.G., Goh, W.C., Dayton, A.I., Lippke, J., Haseltine, W.A. (1986) Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III Nature 319, 555-559.

19. Rosen, C., Sodroski, J.G., Haseltine, W.A. (1985). Location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell 41, 813-823.

20. Sodroski, J.G., Goh, W.C., Rosen, C., Dayton, A., Terwillinger, W., Haseltine, W.A. (1986) A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature 321, 412-417.

21. Jones, K.A., Kadonaga, J.T., Luciw, P.A., Tjian, R. (1986) Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science 232, 755-759.

22. Gorman, G.M., Moffat, L.F., Howard, B.H. (1982). Recombinant genomes which express cloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2, 1044-1051.

23. Fisher, A.G., Collalti, E., Ratner, L., Gallo, R.C., Wong-Staal, F. (1985) A molecular clone of HTLV-III with biological activity. Nature 316, 262-265.

24. Seigel, L.J., Ratner, L., Josephs, S.F., Derse, D., Feinberg, M.B., Reyes, G.R., O'Brien, S.J., Wong-Staal, F. (1986) Transactivation induced by human T-lymphotropic virus type III (HTLV-III) maps to a viral sequence encoding 58 amino acids and lacks tissue specificity. Virology 148, 226-231.

25. Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74, 560-564.

26. Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B., Schaffner, W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41, 521-530.

27. Gottlieb, M.S., Schroff, R., Schanler, H.M., Weisman, J.D., Fan, P.T., Wold, R.A., and Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immuno-deficiency. N. Eng. J. Med. 305, 1426-1431.

28. Barre-Sinoussi, F., Cherman, J.C., Rey, f., Nugeybe, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Brun-Veziet, F., Rouzious, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome (AIDS). Science 220, 868-870.

29. Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224, 497-500.

30. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Refield, R., Oleske, J., Safai, B., White,

31. Kanki, P.J., McLane, M.F., King, N.W., Letvin, N.L., Hunt, R.D., Sehgal, P., Daniel, M.D., Desrosiers, R.C., and Essex, M. (1985). Serologic identification and characterization of a macaque T-lymphotropic retrovirus closely related to human T-lymphotropic retroviruses (HTLV) type III. Science 228, 1199-1201.

32. Kanki, P.J. (1986) New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type ($STLV\text{-}III_{AGM}$). Science 232, 238-242.

33. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M.A., Santos-Ferreria, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J.L., and Montagneir, L. (1986). Isolation of a new human retrovirus from West African patients with AIDS. Science 233, 343-346.

34. Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L., and Alizon, M. (1986). Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nature 324, 691-695.

35. Albert, J., Bredberg, U., Chiodi, F., Bottiger, B., Fenyo, E.M., Norrby, E., and Biberfeld, G. (1987) A new pathogenic human retrovirus of West African origin (SBL-6669) and its relationship to HTLV-IV, LAV-II and HTLV-IIIB. AIDS Research and Human Retroviruses (in press).

36. Dalgleish, A.G., Beverly, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984). The CD4(T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312, 763-767.

37. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J., and Montagnier, L. (1984). T-Lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312, 767-771.

38. Kornfeld, H., Riedel, N., Viglianti, A., Hirsch, V., and Mullins, J.I. (1987). Cloning of HTLV-4 and its relation to simian and human immunodeficiency viruses. Nature 326, 610-613.

39. Franchini, G., Collati, E., Arya, A., Fenyo, E.M., Biberfeld, G., Zagury, J.F., Kanki, P.J., Wong-Staal, F., and Gallo, R.C. (1987). Genetic analysis of a new subgroup of human and simian T-lymphotropic retroviruses: HTLV-IV, LAV-2, SBL6669 and STLV-III$_{AGM}$. AIDS Research and Human Retroviruses vol. 3, no. 1, 11-17.

40. Arya, S.K., Beaver, B., Jagodzinski, L., Ensoli, B., Kanki, P.J., Albert, J., Fenyo, E., Biberfeld, G., Zagury, J.F., Laure, F., Essex, M., Norrby, E., Wong-Staal, F., and Gallo, R.C. (1987). New human and simian HIV related retroviruses possess functional transactivator (tat) gene. Nature (in press).

41. Franchini, G., Gurgo, C., Guo, H.G., Gallo, R.C., Collalti, E., Fargnoli, K.A., Hall, L.F., Wong-Staal, and Reitz, M.S. (1987). Nature 328, 539-543.

42. Hirsch, V., Riedel, N., and Mullins, J.I. (1987). The genome organization of STLV-3 is similar to that of the AIDS virus except for a truncated transmembrane protein. Cell 49, 307-309.

43. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987). Genome organization and transactivation of the human immuno-deficiency virus type 2. Nature 326, 662-669.

44. Zagury, J.F., Franchini, G., Reitz, M., Collalti, E., Starcich, B., Hall, L., Fargnoli, K., Jagodzinski, L., Guo, H.G., Laure, F., Zagury, D., Wong-Staal, F., Gallo, R.C. (1987). The genetic variability between HIV-2 isolates is comparable to the variability among HIV-1. Submitted.

45. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.

46. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

47. Queen, C., and Korn, L.J. (1984). A comprehensive sequence analysis program for the IBM personal computer. Nucleic acid research 12, 581-599.

48. Ratner, L., Starcich, B., Josephs, S., Hahn, B., Reddy, E., Livak, K., Petteway, S., Pearson, M., Haseltine, W., Arya, S.K., and Wong-Staal, F. (1985). Polymorphism of the 3' open reading frame and long terminal repeat sequences of the virus associated with the acquired immunodeficiency syndrome, human T-lymphotropic virus type III. Nucleic Acids Research 13, 219-229.

49. Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA. 80, 3618-3622.

50. diMarzo Veronese, F., Copeland, T.D., DeVico, A.L., Rahman, R., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1986). Characterization of the highly immunogenic p66/p51 as the reverse transcriptase of HTLV-III/LAV. Science 231, 1289-1292.

51. Cease, K.B., Margalit, H., Cornette, J.L., Putney, S.D., Robey, W.G., Ouyang, C., Streicher, H.Z., Fischinger, P.J., Gallo, R.C., DeLisi, C., and Berzofsky, J.A. (1987). Helper T-cell antigenic site identification in the acquired immunodeficiency syndrome virus gp120 envelope protein and induction of immunity in mice to the native protein using a 16-residue synthetic peptide. Proc. Nat. Acad. Sci. USA 84, 4249-4253.

52. Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwilliger, E., and Haseltine, W. (1986). A second post-transcriptional trans-activator gene required for HTLV-III replication. Nature 321, 412-417.

53. Feinberg, M.B., Jarrett, R.F., Aldovini, A., Gallo, R.C., and Wong-Staal, F. (1986). HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. Cell 46, 807-817.

54. Allan, J.S., Coligan, J.E., and McLane, M.F. (1985). A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS patients. Science 230, 810-813.

55. Franchini, G., Robert-Guroff, M., Wong-Staal, F., Ghrayeb, J., Katz, I., Chang, T.W. and Chang, N.T. (1986). Expression of the protein encoded by the 3' open reading frame of HTLV-III in bacteria: Demonstration of its immunoreactivity (?) with human sera. Proc. Natl. Acad. Sci. 83, 5282-5285.

56. Poiesz, B.J., Ruscetti, F.W., Mier, J.W., Woods, A.M., and Gallo, R.C. (1980). T-cell lines established from human T-lymphocyte neoplasias by direct response to T-cell growth factor. Proc. Nat. Acad. Sci. USA 77, 6815-6819.

57. Sanchez-Pescador, R., Power, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W. Renard, A., Randolph, A., Levy, J., Dina, D., and Luciw, P.A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science 227, 484-492.

58. Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A., and Capon, D.J. (1985). Nucleic acid structure and expression of the human AIDS/ lymphadenopathy retrovirus. Nature 313, 450-458.

59. Modrow, S., Hahn, B.H., Shaw, G.M., Gallo, R.C., Wong-Staal, F., and Wolf, H. (1987). Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. 61, 570-578.

60. Gallaher, W.R. (1987). Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell 50, 327-328.

61. Fisher, A.G., Ratner, L., Mitsuya, H., Marselle, L.M., Harper, M.E., Broder, S., Gallo, R.C., and Wong-Staal, F. (1986). Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects. Science 233, 655-659.

62. Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M.D., Desrosiers, R.C., Tiollais, P., and Sonigo, P. (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature 328, 543-547.

63. Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., and Wong-Staal, F. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell 45, 637-648.

64. Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A., and Wain-Hobson, S. (1985). Nucleotide sequence of the Visna lentivirus: Relationship to the AIDS virus. Cell 42, 369-382.

65. Wong-Staal, F., Chanda, P.D., and Ghrayeb, J. (1987). Human T-lymphotropic virus type III: The eight gene. AIDS Research and Human Retroviruses (in press).

66. Aldovini, A., Debouck, C., Rosenberg, M., Arya, S., and Wong-Staal, F. (1987). Expression of the complete transactivator gene product of HTLV-III (AIDS Virus) in E. coli: Its immune reactivity with sera from HTLV-III antibody positive individuals. Proc. Nat. Acad. Sci. USA (in press).

67. Arya, S.K, Guo, C., Joseph, S.F., and Wong-Staal, F. (1985). Transactivator gene of human T-lymphotropic virus type III (HTLV-III). Science 229, 69-73.

68. Clavel, F., Mansinho, K., Chamaret, S., Guetard, D., Favier, V., Nina, J., Santos-Ferreira, M.O., Champalimaud, J.L., and Montagnier, L. (1987). Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. New Eng. J. Md. 316, 1180-1185.

69. Saxinger, W.C., Levine, P.H., Dean, A.G., De-the, G., Moghissi, J., Jaurent, F., Hoh, M., Sarngadharan, M.G., and Gallo, R.C. (1985). Evidence for exposure to HTLV-III in Uganda prior to 1973. Science 227, 1036-1038.

70. Kan, N.C., Franchini, G., Wong-Staal, F., DuBiis, C., Robey, W.G., Lautenberger, J.A., and Papas, T.S. (1986). A novel protein (sor) of HTLV-III expressed in bacteria is immunoreactive with sera from infected individuals. Science 231, 1553-1555.

71. Kanki, P.J., M'Boup, S., Ricard, D., Barin, F., Denis, F., Boye, C., Sangare, L., Travers, K., Albaum, M., Marlink, R., Romet-Lemonne, J.L., and Essex, M. (1987). Human T-lymphotropic virus type 4 and the human immunodeficiency virus in West Africa. Science 236, 827-831.

VI. ACKNOWLEDGEMENTS

We would like to acknowledge the following collaborators in Dr. Robert Gallo's group at the NIH in the work described in this report. The studies on sor, 3'orf and packaging mutations were done in collaboration with Amanda Fisher and Lee Rutner. The LTR and tat activity was done with Steven Josephs and Reza Sedail. The DNA sequencing analysis was done in collaboration with Genoveffa Franchini and Sandra Colombini.

The work on sor mutant function was published by A. Fisher in Science, Vol. 237, 888-893 (1987). The work on the DNA sequencing of the STLV-III_{AGM} and HIV-2_{NIH-Z} has been written up in four papers.

1. Franchini, G., Gurgo, C., Guo H.G., Gallo, R.C., Collatti, E., Fragoli, K.A., Hall, L.F., Wong-Staal, F., Reitz, M.S. (1987) Nature 328, 539-542.

2. Arya, S.K., Beaver, B., Jagodzinski, L., Ensoli, B., Kanki, P.J., Albert, J., Fenyo, E.M., Biberfeld, G., Zagury, J.F., Laure, F., Essex, M., Norrby, E., Wong-Staal, F., Gallo, R.C. (1987). Nature 328, 548-550.

3. Zagury, J.F., Franchini, G., Reitz, M., Collati, E., Stareich, B., Hall, L., Fargnoli, K., Jagodzinski, L., Guo, H.G., Laure, F., Zagury, D., Wong-Staal, F., Gallo, R.C. (1987). The genetic variability between HIV-2 isolates is comparable to the varibility among HIV-1. Submitted.

4. Colombini, S., Arya, S.K., Reitz, M.S., Jagodzinski, L., Beaver, B., Wong-Staal, F. (1987). Structure of the functional form of SIV regulatory genes. Submitted to Science.

DISTRIBUTION LIST FOR DAMD17-86-C-6287

12 copies	Director Walter Reed Army Institute of Research Walter Reed Army Medical Center ATTN: SGRD-UWZ-C Washington, DC 20307-5100
1 copy	Commander US Army Medical Research and Development Command ATTN: SGRD-RMI-S Fort Detrick, Frederick, Maryland 21701-5012
12 copies	Defense Technical Information Center (DTIC) ATTN: DTIC-DDAC Cameron Station Alexandria, VA 22304-6145
1 copy	Dean School of Medicine Uniformed Services University of the Health Sciences 4301 Jones Bridge Road Bethesda, MD 20814-4799
1 copy	Commandant Academy of Health Sciences, US Army ATTN: AHS-CDM Fort Sam, Houston, TX 78234-6100

END

DATE

FILMED

6-88

DTIC